

# LANDWEBER-KACZMARZ METHOD IN BANACH SPACES WITH INEXACT INNER SOLVERS

QINIAN JIN

**ABSTRACT.** In recent years Landweber(-Kaczmarz) method has been proposed for solving nonlinear ill-posed inverse problems in Banach spaces using general convex penalty functions. The implementation of this method involves solving a (nonsmooth) convex minimization problem at each iteration step and the existing theory requires its exact resolution which in general is impossible in practical applications. In this paper we propose a version of Landweber-Kaczmarz method in Banach spaces in which the minimization problem involved in each iteration step is solved inexactly. Based on the  $\varepsilon$ -subdifferential calculus we give a convergence analysis of our method. Furthermore, using Nesterov's strategy, we propose a possible accelerated version of Landweber-Kaczmarz method. Numerical results on computed tomography and parameter identification in partial differential equations are provided to support our theoretical results and to demonstrate our accelerated method.

## 1. Introduction

Regularization of inverse problems has been considered extensively and significant progress has been made; see [10, 18, 21, 26, 34] and references therein. Due to the demand of capturing special features of the reconstructed objects and the demand of dealing with general noise, regularization in Banach spaces has emerged as a highly active research field and many new regularization methods have been proposed and investigated in recent years; one may refer to [5, 17, 19, 23, 24, 25, 33] and the references therein for recent developments.

Because of its simplicity and relatively small complexity per iteration, Landweber iteration and its Kaczmarz version have received extensive attention in inverse problem community [9, 11, 12, 14, 27]. In recent years, several versions of Landweber iteration has been formulated in Banach spaces, see [5, 20, 22, 31]. When solving ill-posed system of the form

$$F_i(x) = y_i, \quad i = 0, \dots, N-1 \quad (1.1)$$

consisting of  $N$  equations, a Kaczmarz version of Landweber iteration in Banach spaces with general uniformly convex penalty functions has been proposed in [22] which cyclically considers each equation in (1.1) in a Gauss-Seidel manner. For these modern versions of Landweber method, each iteration step essentially requires the computation of next iterate  $\xi_+, x_+$  from current iterate  $\xi_c, x_c$  via

$$\begin{cases} \xi_+ = \xi_c - tF'(x_c)^* J(F(x_c) - y), \\ x_+ = \arg \min_x \{\Theta(x) - \langle \xi_+, x \rangle\}, \end{cases}$$

where  $t > 0$  is a step size,  $J$  is a duality mapping,  $F, y$  denote one of  $F_i, y_i$ ,  $F'(x)$  denotes the Fréchet derivative of  $F$ , and  $\Theta$  is a uniformly convex function. Therefore, the implementation of the Landweber(-Kaczmarz) method in Banach spaces

requires to solving a minimization problem of the form

$$x = \arg \min_z \{\Theta(z) - \langle \xi, z \rangle\} \quad (1.2)$$

associated with  $\Theta$  in each iteration step.

The existing convergence theory on Landweber(-Kaczmarz) method in Banach spaces requires the exact resolution of the minimization problem (1.2). For some special  $\Theta$  its exact resolution is possible. However, this minimization problem in general can only be solved inexactly by an iterative procedure which may produce an approximate solution  $\bar{x}$  satisfying

$$\Theta(\bar{x}) - \langle \xi, \bar{x} \rangle \leq \arg \min_z \{\Theta(z) - \langle \xi, z \rangle\} + \varepsilon \quad (1.3)$$

for some small  $\varepsilon > 0$ . Furthermore, numerical simulations indicate that solving (1.2) very accurately in every step does not improve the final reconstruction result but wastes a huge amount of computational time. Therefore, it is necessary to formulate a Landweber-Kaczmarz method with inexact inner solver in each iteration step and to develop the corresponding convergence theory. The inequality (1.3) suggests that the  $\varepsilon$ -subdifferential calculus might be a useful tool for this purpose.

It is well-known that Landweber iteration admits the slow convergence property ([10]) which restricts its applications to wide range of problems. To expand its applied range, it is necessary to introduce some acceleration strategy to fasten its convergence speed. In this paper we will use the Nesterov's strategy in optimization ([29]) to propose an accelerated version of Landweber-Kaczmarz method in Banach spaces in which some extrapolation steps are incorporated. We do not have a theory to guarantee its acceleration effect at this moment, however, we will provide numerical simulations to support the fast convergence property.

This paper is organized as follows. In section 2 we will provide some preliminaries on Banach spaces and convex analysis and derive some useful results concerning  $\varepsilon$ -subdifferential. In section 3 we propose the Landweber-Kaczmarz method with inexact inner solvers, show that it is well-defined, and prove its convergence and regularization property. We then use Nesterov's strategy to propose an accelerated version. We also discuss how to produce the inexact solvers for solving the inner minimization problem at each iteration step of Landweber-Kaczmarz method. Finally, we provide numerical simulations to verify the theoretical results and to demonstrate the fast convergence of the accelerated method.

## 2. Preliminaries

Let  $\mathcal{X}$  be two Banach space whose norm is denoted by  $\|\cdot\|$ . We use  $\mathcal{X}^*$  to denote its dual spaces. For any  $x \in \mathcal{X}$  and  $\xi \in \mathcal{X}^*$ , we write  $\langle \xi, x \rangle = \xi(x)$  for the duality pairing. Let  $\mathcal{Y}$  be another Banach space. By  $\mathcal{L}(\mathcal{X}, \mathcal{Y})$  we denote for the space of all bounded linear operators from  $\mathcal{X}$  to  $\mathcal{Y}$ . For any  $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$  we use  $A^* : \mathcal{Y}^* \rightarrow \mathcal{X}^*$  to denote its adjoint, i.e.

$$\langle A^* \zeta, x \rangle = \langle \zeta, Ax \rangle$$

for any  $x \in \mathcal{X}$  and  $\zeta \in \mathcal{Y}^*$ .

For each  $1 < s < \infty$ , the set-valued mapping  $J_s^\mathcal{X} : \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$  defined by

$$J_s^\mathcal{X}(x) := \{\xi \in \mathcal{X}^* : \|\xi\| = \|x\|^{s-1} \text{ and } \langle \xi, x \rangle = \|x\|^s\}$$

is called the duality mapping of  $\mathcal{X}$  with gauge function  $t \rightarrow t^{s-1}$ . When  $\mathcal{X}$  is uniformly smooth in the sense that its modulus of smoothness

$$\rho_{\mathcal{X}}(t) := \sup\{\|\bar{x} + x\| + \|\bar{x} - x\| - 2 : \|\bar{x}\| = 1, \|x\| \leq t\}$$

satisfies  $\lim_{t \searrow 0} \frac{\rho_{\mathcal{X}}(t)}{t} = 0$ , the duality mapping  $J_s^{\mathcal{X}}$ , for each  $1 < s < \infty$ , is single valued and uniformly continuous on bounded sets.

Given a convex function  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ , we use

$$\mathcal{D}(\Theta) := \{x \in \mathcal{X} : \Theta(x) < \infty\}$$

to denote its effective domain. It is called proper if  $\mathcal{D}(\Theta) \neq \emptyset$ . For a proper convex function  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$  and  $x \in \mathcal{X}$ , we define for any  $\varepsilon \geq 0$  the set

$$\partial_{\varepsilon}\Theta(x) := \{\xi \in \mathcal{X}^* : \Theta(\bar{x}) \geq \Theta(x) + \langle \xi, \bar{x} - x \rangle - \varepsilon \text{ for all } \bar{x} \in \mathcal{X}\}$$

which is called the  $\varepsilon$ -subdifferential of  $\Theta$  at  $x$ . Any element in  $\partial_{\varepsilon}\Theta(x)$  is called an  $\varepsilon$ -subgradient of  $\Theta$  at  $x$ . When  $\varepsilon = 0$ , the  $\varepsilon$ -subdifferential of  $\Theta$  reduces to the subdifferential  $\partial\Theta$ . It is clear that  $\partial_{\varepsilon}\Theta(x) \neq \emptyset$  for some  $\varepsilon \geq 0$  implies  $x \in \mathcal{D}(\Theta)$ . If  $\Theta$  is lower semi-continuous, then for any  $x \in \mathcal{D}(\Theta)$ , the  $\varepsilon$ -subdifferential  $\partial_{\varepsilon}\Theta(x)$  is always non-empty for any  $\varepsilon > 0$ , see [35, Theorem 2.4.4]; however,  $\partial\Theta(x)$  can be empty in general.

For  $\xi \in \partial_{\varepsilon}\Theta(x)$  with  $\varepsilon \geq 0$ , we may introduce

$$D_{\xi}^{\varepsilon}\Theta(\bar{x}, x) := \Theta(\bar{x}) - \Theta(x) - \langle \xi, \bar{x} - x \rangle + \varepsilon, \quad \forall \bar{x} \in \mathcal{X}$$

which is called the  $\varepsilon$ -Bregman distance induced by  $\Theta$  at  $x$  in the direction  $\xi$ . It is clear that

$$D_{\xi}^{\varepsilon}\Theta(\bar{x}, x) \geq 0.$$

When  $\varepsilon = 0$ , the  $\varepsilon$ -Bregman distance becomes the well-known Bregman distance [7] which will be denoted by  $D_{\xi}\Theta(\bar{x}, x)$ . It should be pointed out that  $\varepsilon$ -Bregman distance is not a metric distance in general. Nevertheless, as the following result shows, the  $\varepsilon$ -Bregman distance can be used to detect information under the norm if  $\Theta$  is  $p$ -convex for some  $p \geq 2$  in the sense that there is a constant  $c_0 > 0$  such that

$$\Theta(t\bar{x} + (1-t)x) + c_0 t(1-t)\|\bar{x} - x\|^p \leq t\Theta(\bar{x}) + (1-t)\Theta(x) \quad (2.1)$$

for all  $0 \leq t \leq 1$  and  $\bar{x}, x \in \mathcal{X}$ .

**Lemma 2.1.** *Let  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$  be a proper, lower semi-continuous function that is  $p$ -convex in the sense of (2.1). If  $\xi \in \partial_{\varepsilon}\Theta(x)$  for some  $\varepsilon \geq 0$ , then*

$$c_0\|\bar{x} - x\|^p \leq 2D_{\xi}^{\varepsilon}\Theta(\bar{x}, x) + 2\varepsilon. \quad (2.2)$$

for any  $\bar{x} \in \mathcal{X}$ .

*Proof.* Since  $\Theta$  is  $p$ -convex and  $\xi \in \partial_{\varepsilon}\Theta(x)$ , we have for any  $0 < t < 1$  that

$$\begin{aligned} c_0 t(1-t)\|\bar{x} - x\|^p &\leq t\Theta(x) + (1-t)\Theta(\bar{x}) - \Theta(tx + (1-t)\bar{x}) \\ &\leq t\Theta(x) + (1-t)\Theta(\bar{x}) - [\Theta(x) + (1-t)\langle \xi, \bar{x} - x \rangle - \varepsilon] \\ &= (1-t)[\Theta(\bar{x}) - \Theta(x) - \langle \xi, \bar{x} - x \rangle + \varepsilon] + t\varepsilon \\ &= (1-t)D_{\xi}^{\varepsilon}\Theta(\bar{x}, x) + t\varepsilon. \end{aligned}$$

By taking  $t = 1/2$  we then obtain (2.2).  $\square$

In convex analysis, the Legendre-Fenchel conjugate is an important notion. Given a proper convex function  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ , its Legendre-Fenchel conjugate is defined by

$$\Theta^*(\xi) := \sup_{x \in \mathcal{X}} \{ \langle \xi, x \rangle - \Theta(x) \}, \quad \forall \xi \in \mathcal{X}^*.$$

As an immediate consequence of the definition, one can see, for any  $\varepsilon \geq 0$ , that

$$\xi \in \partial_\varepsilon \Theta(x) \iff \Theta(x) + \Theta^*(\xi) \leq \langle \xi, x \rangle + \varepsilon. \quad (2.3)$$

If, in addition,  $\Theta$  is lower semi-continuous, then ([35, Theorem 2.4.4])

$$\xi \in \partial_\varepsilon \Theta(x) \iff x \in \partial_\varepsilon \Theta^*(\xi). \quad (2.4)$$

For a proper, lower semi-continuous,  $p$ -convex function, even if it is non-smooth, its Legendre-Fenchel conjugate can have enough regularity as the following result indicates.

**Lemma 2.2.** *Let  $\mathcal{X}$  be a reflexive Banach space and let  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$  be a proper, lower semi-continuous function that is  $p$ -convex in the sense of (2.1). Then  $\mathcal{D}(\Theta^*) = \mathcal{X}^*$ ,  $\Theta^*$  is Fréchet differentiable, and its gradient  $\nabla \Theta^* : \mathcal{X}^* \rightarrow \mathcal{X}$  satisfies*

$$\|\nabla \Theta^*(\xi) - \nabla \Theta^*(\eta)\| \leq \left( \frac{\|\xi - \eta\|}{2c_0} \right)^{\frac{1}{p-1}} \quad (2.5)$$

which consequently implies

$$|\Theta^*(\eta) - \Theta^*(\xi) - \langle \eta - \xi, \nabla \Theta^*(\xi) \rangle| \leq \frac{1}{p^*(2c_0)^{p^*-1}} \|\xi - \eta\|^{p^*} \quad (2.6)$$

for any  $\xi, \eta \in \mathcal{X}^*$ , where  $p^*$  is the number conjugate to  $p$ , i.e.  $1/p + 1/p^* = 1$ .

*Proof.* See [35, Theorem 3.5.10 and Corollary 3.5.11].  $\square$

Finally we conclude this section by providing a result which show that, when  $\Theta$  is  $p$ -convex, then, for any  $x \in \partial_\varepsilon \Theta^*(\xi)$ , the distance from  $x$  to  $\nabla \Theta^*(\xi)$  can be controlled in terms of  $\varepsilon$ .

**Lemma 2.3.** *Let  $\mathcal{X}$  be a reflexive Banach space and let  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$  be a proper, lower semi-continuous function that is  $p$ -convex in the sense of (2.1). If  $x \in \mathcal{X}$  and  $\xi \in \mathcal{X}^*$  satisfy  $\xi \in \partial_\varepsilon \Theta(x)$  for some  $\varepsilon \geq 0$ , then for any  $\eta \in \mathcal{X}^*$  there holds*

$$\langle \eta, x - \nabla \Theta^*(\xi) \rangle \leq \varepsilon + \frac{1}{p^*(2c_0)^{p^*-1}} \|\eta\|^{p^*} \quad (2.7)$$

and hence

$$\|x - \nabla \Theta^*(\xi)\|^p \leq \frac{p}{2c_0} \varepsilon. \quad (2.8)$$

*Proof.* Since  $\xi \in \partial_\varepsilon \Theta(x)$ , by (2.4) we have  $x \in \partial_\varepsilon \Theta^*(\xi)$  and hence

$$\Theta^*(\xi + \eta) \geq \Theta^*(\xi) + \langle \eta, x \rangle - \varepsilon.$$

By using (2.6) in Lemma 2.2, we also have

$$\Theta^*(\xi + \eta) \leq \Theta^*(\xi) + \langle \eta, \nabla \Theta^*(\xi) \rangle + \frac{1}{p^*(2c_0)^{p^*-1}} \|\eta\|^{p^*}.$$

Combining the above two inequalities we obtain

$$\langle \eta, x \rangle - \varepsilon \leq \langle \eta, \nabla \Theta^*(\xi) \rangle + \frac{1}{p^*(2c_0)^{p^*-1}} \|\eta\|^{p^*}$$

which shows (2.7). By taking  $\eta \in 2c_0 J_p^\mathcal{X}(x - \nabla\Theta^*(\xi))$  in (2.7) and using the properties of  $J_p^\mathcal{X}$ , we then obtain

$$2c_0 \|x - \nabla\Theta^*(\xi)\|^p \leq \varepsilon + \frac{2c_0}{p^*} \|x - \nabla\Theta^*(\xi)\|^p$$

which shows (2.8).  $\square$

### 3. The method

We consider the system

$$F_i(x) = y_i, \quad i = 0, \dots, N-1 \quad (3.1)$$

consisting of  $N$  equations, where, for each  $i = 0, \dots, N-1$ ,  $F_i : \mathcal{D}(F_i) \subset \mathcal{X} \rightarrow \mathcal{Y}_i$  is an operator between two reflexive Banach spaces  $\mathcal{X}$  and  $\mathcal{Y}_i$ . Such systems arise in many practical applications including various tomography problems using multiple exterior measurements.

We will assume that (3.1) has a solution which consequently implies that

$$\mathcal{D} := \bigcap_{i=0}^{N-1} \mathcal{D}(F_i) \neq \emptyset.$$

In practical applications, instead of  $y_i$  we only have noisy data  $y_i^\delta$  satisfying

$$\|y_i^\delta - y_i\| \leq \delta, \quad i = 0, \dots, N-1 \quad (3.2)$$

with a small noise level  $\delta > 0$ . How to use  $y_i^\delta$  to produce an approximate solution of (3.1) is an important question. In [22] we proposed a Landweber iteration of Kaczmarz type which makes use of every equation in (3.1) cyclically. In order to capture the feature of the sought solution, general convex functions  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$  have been used in [22] as penalty terms.

We will make the following assumption, where  $B_\rho(x_0) := \{x \in \mathcal{X} : \|x - x_0\| \leq \rho\}$ .

**Assumption 3.1.** (a)  $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$  is proper, lower semi-continuous and  $p$ -convex in the sense of (2.1).

(b) There exist  $\rho > 0$ ,  $x_0 \in \mathcal{X}$  and  $\xi_0 \in \partial\Theta(x_0)$  such that  $B_{2\rho}(x_0) \subset \mathcal{D}$  and (3.1) has a solution  $x^\dagger \in \mathcal{D}(\Theta)$  with

$$D_{\xi_0}\Theta(x^\dagger, x_0) \leq \frac{1}{4}c_0\rho^p; \quad (3.3)$$

(c) For each  $i = 0, \dots, N-1$  there exists  $\{L_i(x) : \mathcal{X} \rightarrow \mathcal{Y}_i\}_{x \in B_{2\rho}(x_0)} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y}_i)$  such that  $x \rightarrow L_i(x)$  is continuous on  $B_{2\rho}(x_0)$  and there is  $0 \leq \gamma < 1$  such that

$$\|F_i(\bar{x}) - F_i(x) - L_i(x)(\bar{x} - x)\| \leq \gamma \|F_i(\bar{x}) - F_i(x)\|$$

for all  $\bar{x}, x \in B_{2\rho}(x_0)$ .

According to Assumption 3.1 (c), we can find a constant  $B > 0$  such that

$$\|L_i(x)\| \leq B \quad \forall x \in B_{2\rho}(x_0) \text{ and } i = 0, \dots, N-1. \quad (3.4)$$

Moreover

$$\|F_i(\bar{x}) - F_i(x)\| \leq \frac{1}{1-\gamma} \|L_i(x)(\bar{x} - x)\| \leq \frac{B}{1-\gamma} \|\bar{x} - x\| \quad (3.5)$$

for all  $\bar{x}, x \in B_{2\rho}(x_0)$  which shows that  $F_i$  is continuous on  $B_{2\rho}(x_0)$  for each  $i = 0, \dots, N-1$ .

The formulation of the Landweber iteration of Kaczmarz type in [22] involves in each iteration step a minimization problem of the form

$$x := \arg \min_{z \in \mathcal{X}} \{\Theta(z) - \langle \xi, z \rangle\} \quad (3.6)$$

for any  $\xi \in \mathcal{X}^*$ . The convergence result developed there requires to solving (3.6) exactly. The exact solution of (3.6) can be found for some special  $\Theta$ . However, this minimization problem in general can only be solved inexactly by iterative procedures. Furthermore, numerical simulations indicate that solving (3.6) very accurately in every step does not improve the final reconstruction result. Therefore, it is necessary to formulate a Landweber-Kaczmarz method with inexact inner solver in each iteration step and to develop the corresponding convergence theory.

Concerning the inexact resolution of (3.6), we make the following assumption.

**Assumption 3.2.** *For any given  $\varepsilon > 0$  there is a procedure  $S_\varepsilon : \mathcal{X}^* \rightarrow \mathcal{X}$  for solving (3.6) such that for any  $\xi \in \mathcal{X}^*$ , the element  $x := S_\varepsilon(\xi)$  satisfies*

$$\Theta(x) - \langle \xi, x \rangle \leq \min_{z \in \mathcal{X}} \{\Theta(z) - \langle \xi, z \rangle\} + \varepsilon. \quad (3.7)$$

Moreover, for each  $\varepsilon > 0$ , the mapping  $S_\varepsilon : \mathcal{X}^* \rightarrow \mathcal{X}$  is continuous.

In subsection 3.5 we will discuss how to produce the inexact procedure  $S_\varepsilon$  by using concrete examples of  $\Theta$  including the total variation like convex penalty functions.

**3.1. The method with noisy data.** We are ready to formulate our Landweber-Kaczmarz method with inexact inner solvers. We will take  $1 < s < \infty$  and let  $J_s^{\mathcal{Y}_i}$  denote the duality mapping over  $\mathcal{Y}_i$  with gauge function  $t \rightarrow t^{s-1}$ . Given an integer  $n$ , we set  $i_n := n \pmod{N}$ .

**Algorithm 3.1** (Landweber-Kaczmarz method with noisy data).

Let  $\beta_0 > 0$ ,  $\beta_1 > 0$  and  $\tau > 1$  be suitably chosen numbers, and let  $\{\varepsilon_n\}_{n \geq 0}$  be a sequence of positive numbers satisfying  $\sum_{n=0}^{\infty} \varepsilon_n < \infty$ .

- (i) Pick  $x_0 \in \mathcal{X}$  and  $\xi_0 \in \mathcal{X}^*$  such that  $\xi_0 \in \partial\Theta(x_0)$ .
- (ii) Let  $\xi_0^\delta := \xi_0$  and  $x_0^\delta := x_0$ . Let  $q_{-1} = 0$  and let  $\sigma > 0$  be a small number. For  $n \geq 0$  we define  $r_n^\delta = F_{i_n}(x_n^\delta) - y_{i_n}^\delta$  and

$$q_n = \begin{cases} q_{n-1} + 1 & \text{if } \|r_n^\delta\|^p + \sigma\varepsilon_n \leq (\tau\delta)^p, \\ 0 & \text{otherwise.} \end{cases}$$

We then update

$$\begin{cases} \xi_{n+1}^\delta = \xi_n^\delta - \mu_n^\delta L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta), \\ x_{n+1}^\delta = S_{\varepsilon_{n+1}}(\xi_{n+1}^\delta), \end{cases} \quad (3.8)$$

where

$$\mu_n^\delta = \begin{cases} \tilde{\mu}_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{1-\frac{s}{p}} & \text{if } \|r_n^\delta\|^p + \sigma\varepsilon_n > (\tau\delta)^p, \\ 0 & \text{otherwise} \end{cases} \quad (3.9)$$

with

$$\tilde{\mu}_n^\delta = \min \left\{ \frac{\beta_0 \|r_n^\delta\|^{p(s-1)}}{\|L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta)\|^p}, \beta_1 \right\}.$$

- (iii) Let  $n_\delta$  be the first integer such that  $q_{n_\delta} = N$  and use  $x_{n_\delta}^\delta$  as an approximate solution.

In Algorithm 3.1, each  $\xi_n^\delta$  is determined by  $F_{i_n}$  completely without involving  $\Theta$ , and each  $x_n^\delta$  is defined by the inexact procedure specified in Assumption 3.2 for solving (3.6) which is independent of  $F_i$ ,  $i = 0, \dots, N-1$ . This splitting character can make the implementation of Algorithm 3.1 efficiently. Furthermore, the definition of  $x_n^\delta$  implies that

$$\Theta(x_n^\delta) - \langle \xi_n^\delta, x_n^\delta \rangle \leq \Theta(x) - \langle \xi_n^\delta, x \rangle + \varepsilon_n, \quad \forall x \in \mathcal{X}$$

which shows that

$$\xi_n^\delta \in \partial_{\varepsilon_n} \Theta(x_n^\delta). \quad (3.10)$$

We will use this fact in the forthcoming convergence analysis.

We first prove the following basic result which shows that Algorithm 3.1 is well-defined.

**Lemma 3.1.** *Let  $\mathcal{X}$  and  $\mathcal{Y}_i$  be reflexive Banach spaces with  $\mathcal{Y}_i$  being uniformly smooth, let  $\Theta$  and  $F_i$ ,  $i = 0, \dots, N-1$  satisfy Assumption 3.1, and let  $\{\varepsilon_n\}_{n \geq 0}$  be a sequence of positive numbers satisfying*

$$16 \sum_{n=0}^{\infty} \varepsilon_n \leq c_0 \rho^p. \quad (3.11)$$

Let  $\beta > 1$  be a constant such that  $\beta\gamma < 1$ . If  $\beta_0 > 0$  and  $\tau > 1$  are chosen such that

$$c_1 := \frac{1}{\beta} - \gamma - \frac{1+\gamma}{\tau} - \frac{2}{p^*} \left( \frac{\beta_0}{2c_0} \right)^{p^*-1} > 0 \quad (3.12)$$

and if  $\sigma > 0$  is chosen such that  $\kappa\beta_1\sigma \leq 1$ , where

$$\kappa = \begin{cases} 1 & \text{if } p \geq s \\ (\beta^{\frac{p}{s-p}} - 1)^{\frac{p-s}{p}} & \text{if } p < s, \end{cases} \quad (3.13)$$

then for Algorithm 3.1 there hold

- (i)  $x_n^\delta \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  for all  $n = 0, 1, \dots$ ;
- (ii) the method terminates after  $n_\delta < \infty$  iteration steps;
- (iii) for any solution  $\hat{x}$  of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  there holds

$$D_{\xi_{n+1}^\delta}^{\varepsilon_{n+1}} \Theta(\hat{x}, x_{n+1}^\delta) - D_{\xi_n^\delta}^{\varepsilon_n} \Theta(\hat{x}, x_n^\delta) \leq 2\varepsilon_n + \varepsilon_{n+1} \quad (3.14)$$

for all  $n \geq 0$ . Here we may take  $\varepsilon_0 = 0$  because  $\xi_0 \in \partial\Theta(x_0)$ .

*Proof.* Let  $\hat{x}$  be any solution of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ . We first show that, if  $x_n^\delta \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  for some  $n \geq 0$ , then

$$D_{\xi_{n+1}^\delta}^{\varepsilon_{n+1}} \Theta(\hat{x}, x_{n+1}^\delta) - D_{\xi_n^\delta}^{\varepsilon_n} \Theta(\hat{x}, x_n^\delta) \leq -c_1 \mu_n^\delta (\|r_n^\delta\|^p + \sigma \varepsilon_n)^{\frac{s}{p}} + 2\varepsilon_n + \varepsilon_{n+1}. \quad (3.15)$$

To see this, we consider

$$\Delta_n := D_{\xi_{n+1}^\delta}^{\varepsilon_{n+1}} \Theta(\hat{x}, x_{n+1}^\delta) - D_{\xi_n^\delta}^{\varepsilon_n} \Theta(\hat{x}, x_n^\delta)$$

which can be written as

$$\begin{aligned} \Delta_n &= [\Theta(x_n^\delta) - \langle \xi_n^\delta, x_n^\delta \rangle - \varepsilon_n] + [\langle \xi_{n+1}^\delta, x_{n+1}^\delta \rangle - \Theta(x_{n+1}^\delta)] \\ &\quad - \langle \xi_{n+1}^\delta - \xi_n^\delta, \hat{x} \rangle + \varepsilon_{n+1}. \end{aligned}$$

Since  $\xi_n^\delta \in \partial_{\varepsilon_n} \Theta(x_n^\delta)$ , we have from (2.3) that

$$\Theta(x_n^\delta) - \langle \xi_n^\delta, x_n^\delta \rangle - \varepsilon_n \leq -\Theta^*(\xi_n^\delta).$$

By the definition of  $\Theta^*$  we also have

$$\langle \xi_{n+1}^\delta, x_{n+1}^\delta \rangle - \Theta(x_{n+1}^\delta) \leq \Theta^*(\xi_{n+1}^\delta).$$

Therefore

$$\begin{aligned} \Delta_n &\leq \Theta^*(\xi_{n+1}^\delta) - \Theta^*(\xi_n^\delta) - \langle \xi_{n+1}^\delta - \xi_n^\delta, \hat{x} \rangle + \varepsilon_{n+1} \\ &= \Theta^*(\xi_{n+1}^\delta) - \Theta^*(\xi_n^\delta) - \langle \xi_{n+1}^\delta - \xi_n^\delta, \nabla \Theta^*(\xi_n^\delta) \rangle \\ &\quad + \langle \xi_{n+1}^\delta - \xi_n^\delta, \nabla \Theta^*(\xi_n^\delta) - x_n^\delta \rangle \\ &\quad + \langle \xi_{n+1}^\delta - \xi_n^\delta, x_n^\delta - \hat{x} \rangle + \varepsilon_{n+1}. \end{aligned}$$

Because  $\Theta$  is  $p$ -convex, we may use (2.6) in Lemma 2.2 and the definition of  $\xi_{n+1}^\delta$  to obtain

$$\begin{aligned} \Delta_n &\leq \frac{1}{p^*(2c_0)^{p^*-1}} \|\xi_{n+1}^\delta - \xi_n^\delta\|^{p^*} + \langle \xi_{n+1}^\delta - \xi_n^\delta, \nabla \Theta^*(\xi_n^\delta) - x_n^\delta \rangle \\ &\quad + \mu_n^\delta \langle J_s^{\mathcal{Y}_{i_n}}(r_n^\delta), L_{i_n}(x_n^\delta)(\hat{x} - x_n^\delta) \rangle + \varepsilon_{n+1}. \end{aligned} \quad (3.16)$$

Since  $\xi_n^\delta \in \partial_{\varepsilon_n} \Theta(x_n^\delta)$ , we may use Lemma 2.3 to derive that

$$\langle \xi_{n+1}^\delta - \xi_n^\delta, \nabla \Theta^*(\xi_n^\delta) - x_n^\delta \rangle \leq \varepsilon_n + \frac{1}{p^*(2c_0)^{p^*-1}} \|\xi_{n+1}^\delta - \xi_n^\delta\|^{p^*}.$$

Plugging this estimate into (3.16) and using the definition of  $\xi_{n+1}^\delta$  it follows that

$$\begin{aligned} \Delta_n &\leq \frac{2}{p^*(2c_0)^{p^*-1}} (\mu_n^\delta)^{p^*} \|L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta)\|^{p^*} + \varepsilon_n + \varepsilon_{n+1} \\ &\quad + \mu_n^\delta \langle J_s^{\mathcal{Y}_{i_n}}(r_n^\delta), L_{i_n}(x_n^\delta)(\hat{x} - x_n^\delta) \rangle. \end{aligned}$$

By writing

$$L_{i_n}(x_n^\delta)(\hat{x} - x_n^\delta) = -r_n^\delta - [y_{i_n}^\delta - F_{i_n}(x_n^\delta) - L_{i_n}(x_n^\delta)(\hat{x} - x_n^\delta)],$$

we may use the condition  $\|y_{i_n}^\delta - y_{i_n}\| \leq \delta$ , Assumption 3.1 (c), and the properties of  $J_s^{\mathcal{Y}_{i_n}}$  to obtain

$$\begin{aligned} &\langle J_s^{\mathcal{Y}_{i_n}}(r_n^\delta), L_{i_n}(x_n^\delta)(\hat{x} - x_n^\delta) \rangle \\ &\leq -\|r_n^\delta\|^s + \|r_n^\delta\|^{s-1} \|y_{i_n}^\delta - F_{i_n}(x_n^\delta) - L_{i_n}(x_n^\delta)(\hat{x} - x_n^\delta)\| \\ &\leq -\|r_n^\delta\|^s + \|r_n^\delta\|^{s-1} ((1 + \gamma)\delta + \gamma\|r_n^\delta\|). \end{aligned}$$

Therefore

$$\begin{aligned} \Delta_n &\leq \frac{2}{p^*(2c_0)^{p^*-1}} (\mu_n^\delta)^{p^*} \|L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta)\|^{p^*} + \varepsilon_n + \varepsilon_{n+1} \\ &\quad - \mu_n^\delta \|r_n^\delta\|^s + \mu_n^\delta \|r_n^\delta\|^{s-1} ((1 + \gamma)\delta + \gamma\|r_n^\delta\|). \end{aligned} \quad (3.17)$$

By the definition of  $\mu_n^\delta$  we can see that

$$\begin{aligned} (\mu_n^\delta)^{p^*} \|L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta)\|^{p^*} &= \mu_n^\delta (\mu_n^\delta)^{p^*-1} \|L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta)\|^{p^*} \\ &\leq \beta_0^{p^*-1} \mu_n^\delta \|r_n^\delta\|^{(s-1)p^*} (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{(p-s)(p^*-1)}{p}} \\ &\leq \beta_0^{p^*-1} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} \end{aligned}$$



and

$$\begin{aligned} \mu_n^\delta \|r_n^\delta\|^{s-1} ((1+\gamma)\delta + \gamma\|r_n^\delta\|) &\leq \frac{1+\gamma}{\tau} \mu_n^\delta \|r_n^\delta\|^{s-1} (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{1}{p}} + \gamma\mu_n^\delta \|r_n^\delta\|^s \\ &\leq \left(\frac{1+\gamma}{\tau} + \gamma\right) \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}}. \end{aligned}$$

Combining the above two estimates with (3.17) we can obtain

$$\begin{aligned} \Delta_n &\leq \left[ \frac{2}{p^*} \left( \frac{\beta_0}{2c_0} \right)^{p^*-1} + \frac{1+\gamma}{\tau} + \gamma \right] \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} \\ &\quad - \mu_n^\delta \|r_n^\delta\|^s + \varepsilon_n + \varepsilon_{n+1}. \end{aligned} \quad (3.18)$$

We next consider the term  $\mu_n^\delta \|r_n^\delta\|^s$ . We claim that

$$\mu_n^\delta \|r_n^\delta\|^s \geq \frac{1}{\beta} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} - \kappa \tilde{\mu}_n^\delta \sigma\varepsilon_n, \quad (3.19)$$

where  $\kappa > 0$  is the constant defined by (3.13). Indeed, this is trivial when  $\mu_n^\delta = 0$ . We only need to consider the case that  $\mu_n^\delta \neq 0$ . If  $p \geq s$ , then

$$\begin{aligned} \mu_n^\delta \|r_n^\delta\|^s &= \tilde{\mu}_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{p-s}{p}} \|r_n^\delta\|^s \geq \tilde{\mu}_n^\delta \|r_n^\delta\|^p \\ &= \tilde{\mu}_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n) - \tilde{\mu}_n^\delta \sigma\varepsilon_n \\ &= \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} - \tilde{\mu}_n^\delta \sigma\varepsilon_n. \end{aligned}$$

If  $p < s$ , we may use the inequality  $(a+b)^t \leq \beta a^t + \beta(\beta^{\frac{1}{t-1}} - 1)^{1-t} b^t$  for  $a, b \geq 0$  and  $t > 1$  to derive that

$$\|r_n^\delta\|^s \geq \frac{1}{\beta} (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} - (\beta^{\frac{p}{s-p}} - 1)^{\frac{p-s}{p}} (\sigma\varepsilon_n)^{\frac{s}{p}}.$$

Thus, by using  $\mu_n^\delta = \tilde{\mu}_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{p-s}{p}} \leq \tilde{\mu}_n^\delta (\sigma\varepsilon_n)^{\frac{p-s}{p}}$ , we have

$$\mu_n^\delta \|r_n^\delta\|^s \geq \frac{1}{\beta} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} - (\beta^{\frac{p}{s-p}} - 1)^{\frac{p-s}{p}} \tilde{\mu}_n^\delta \sigma\varepsilon_n.$$

We therefore obtain (3.19).

Combining (3.18) and (3.19) we thus have

$$\Delta_n \leq -c_1 \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} + \kappa \tilde{\mu}_n^\delta \sigma\varepsilon_n + \varepsilon_n + \varepsilon_{n+1},$$

where  $c_1 > 0$  is the constant defined by (3.12). Since  $\tilde{\mu}_n^\delta \leq \beta_1$  and  $\kappa\beta_1\sigma \leq 1$ , we therefore obtain (3.15).

Now we use an induction argument to show that  $x_n \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  for all  $n \geq 0$ . This is trivial when  $n = 0$ . Assume that there is some  $m \geq 0$  such that  $x_n \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  for  $0 \leq n < m$ . Thus (3.15) holds for all  $0 \leq n < m$ . By taking  $\hat{x} = x^\dagger$  in (3.15) and summing it over these  $n$  gives

$$D_{\xi_m^\delta}^\varepsilon \Theta(x^\dagger, x_m) - D_{\xi_0^\delta} \Theta(x^\dagger, x_0) \leq \sum_{n=0}^{m-1} (2\varepsilon_n + \varepsilon_{n+1}).$$

In view of Lemma 2.1, (3.3) in Assumption 3.1 and (3.11) we can obtain

$$\begin{aligned} c_0 \|x_m - x^\dagger\|^p &\leq 2D_{\xi_m^{\varepsilon_m}}^{\varepsilon_m} \Theta(x^\dagger, x_m^\delta) + 2\varepsilon_m \leq 2D_{\xi_0} \Theta(x^\dagger, x_0) + 8 \sum_{n=0}^{\infty} \varepsilon_n \\ &\leq \frac{1}{2} c_0 \rho^p + \frac{1}{2} c_0 \rho^p = c_0 \rho^p. \end{aligned}$$

This implies that  $\|x_m^\delta - x^\dagger\| \leq \rho$ . By virtue of (3.3) and Lemma 2.1 we also have  $\|x_0 - x^\dagger\| \leq \rho$ . Therefore  $\|x_m^\delta - x_0\| \leq 2\rho$ , i.e.  $x_m^\delta \in B_{2\rho}(x_0)$ . Consequently, (3.15) holds for all  $n \geq 0$  which gives (3.14) immediately. By summing (3.15) over  $n$  from 0 to  $\infty$  we can obtain

$$c_1 \sum_{n=0}^{\infty} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} \leq D_{\xi_0} \Theta(\hat{x}, x_0) + 3 \sum_{n=0}^{\infty} \varepsilon_n < \infty. \quad (3.20)$$

Finally we show that  $n_\delta < \infty$ . If it is not true, then for each integer  $d \geq 0$  there is at least one integer  $k$  with  $Nd \leq k \leq Nd + N - 1$  such that  $\|r_k^\delta\|^p + \sigma\varepsilon_k > (\tau\delta)^p$ . Consequently

$$\mu_k^\delta = \tilde{\mu}_k^\delta (\|r_k^\delta\|^p + \sigma\varepsilon_k)^{1-\frac{s}{p}} \geq c_2 (\|r_k^\delta\|^p + \sigma\varepsilon_k)^{1-\frac{s}{p}}$$

with  $c_2 = \min\{\beta_0 B^{-p}, \beta_1\}$ , where we used (3.4). Therefore

$$\sum_{n=Nd}^{Nd+N-1} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} \geq \mu_k^\delta (\|r_k^\delta\|^p + \sigma\varepsilon_k)^{\frac{s}{p}} \geq c_2 (\|r_k^\delta\|^p + \sigma\varepsilon_k) \geq c_2 (\tau\delta)^p$$

and hence

$$\sum_{n=0}^{\infty} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} = \sum_{d=0}^{\infty} \sum_{n=Nd}^{Nd+N-1} \mu_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{\frac{s}{p}} \geq c_2 \sum_{d=0}^{\infty} (\tau\delta)^p = \infty$$

which is a contradiction to (3.20).  $\square$

**3.2. The method with exact data.** In order to prove the regularization property of Algorithm 3.1, we first consider its counterpart where the noisy data  $y_i^\delta$  are replaced by the exact data  $y_i$ . That is, we will consider the following algorithm.

**Algorithm 3.2** (Landweber-Kaczmarz method with exact data).

Let  $\beta_0 > 0$ ,  $\beta_1 > 0$ ,  $\sigma > 0$ ,  $\{\varepsilon_n\}$ ,  $x_0 \in \mathcal{X}$  and  $\xi_0 \in \mathcal{X}^*$  be the same as in Algorithm 3.1. For  $n = 0, 1, \dots$  we define  $r_n = F_{i_n}(x_n) - y_{i_n}$  and update

$$\begin{aligned} \xi_{n+1} &= \xi_n - \mu_n L_{i_n}(x_n)^* J_s^{\mathcal{Y}_{i_n}}(r_n), \\ x_{n+1} &= S_{\varepsilon_{n+1}}(\xi_{n+1}), \end{aligned} \quad (3.21)$$

where  $i_n = n \pmod{N}$  and

$$\mu_n = \tilde{\mu}_n (\|r_n\|^p + \sigma\varepsilon_n)^{1-\frac{s}{p}} \quad (3.22)$$

with

$$\tilde{\mu}_n = \min \left\{ \frac{\beta_0 \|r_n\|^{p(s-1)}}{\|L_{i_n}(x_n)^* J_s^{\mathcal{Y}_{i_n}}(r_n)\|^p}, \beta_1 \right\}.$$

By using the same argument in the proof of Lemma 3.1, we can obtain the following result on Algorithm 3.2.

**Lemma 3.2.** *Let  $\mathcal{X}$  and  $\mathcal{Y}_i$  be reflexive Banach spaces with  $\mathcal{Y}_i$  being uniformly smooth, let  $\Theta$  and  $F_i$ ,  $i = 0, \dots, N-1$  satisfy Assumption 3.1, and let  $\{\varepsilon_n\}$  be a sequence of positive numbers satisfying (3.11). Let  $\beta > 1$  be a constant such that  $\beta\gamma < 1$ . If  $\beta_0 > 0$  is chosen such that*

$$\frac{1}{\beta} - \gamma - \frac{2}{p^*} \left( \frac{\beta_0}{2c_0} \right)^{p^*-1} > 0 \quad (3.23)$$

*and if  $\sigma > 0$  is chosen such that  $\kappa\beta_1\sigma \leq 1$ , where  $\kappa$  is defined by (3.13), then for Algorithm 3.2 there holds  $x_n \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  for all  $n = 0, 1, \dots$ , and for any solution  $\hat{x}$  of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  there hold*

$$D_{\xi_{n+1}}^{\varepsilon_{n+1}}\Theta(\hat{x}, x_{n+1}) - D_{\xi_n}^{\varepsilon_n}\Theta(\hat{x}, x_n) \leq 2\varepsilon_n + \varepsilon_{n+1}, \quad n = 0, 1, \dots \quad (3.24)$$

$$\sum_{n=0}^{\infty} \mu_n (\|r_n\|^p + \sigma\varepsilon_n)^{\frac{\beta}{p}} < \infty. \quad (3.25)$$

By the definition of  $\mu_n$  and (3.4) one can see that

$$\mu_n (\|r_n\|^p + \sigma\varepsilon_n)^{\frac{\beta}{p}} \geq \min\{\beta_0 B^{-p}, \beta_1\} (\|r_n\|^p + \sigma\varepsilon_n).$$

Thus, it follows from (3.25) that

$$\sum_{n=0}^{\infty} (\|r_n\|^p + \varepsilon_n) < \infty. \quad (3.26)$$

In the following we will show that the sequence  $\{x_n\}$  defined by Algorithm 3.2 converges to a solution of (3.1). We first prove that the sequence  $\{D_{\xi_n}^{\varepsilon_n}\Theta(\hat{x}, x_n)\}$  is convergent.

**Lemma 3.3.** *Let all the conditions in Lemma 3.2 hold. Then, for any solution  $\hat{x}$  of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ , the sequence  $\{D_{\xi_n}^{\varepsilon_n}\Theta(\hat{x}, x_n)\}$  is convergent.*

*Proof.* Let

$$a_n := D_{\xi_n}^{\varepsilon_n}\Theta(\hat{x}, x_n) \quad \text{and} \quad \beta_n := 2\varepsilon_n + \varepsilon_{n+1}.$$

It follows from (3.24) in Lemma 3.2 and  $\sum_n \varepsilon_n < \infty$  that

$$0 \leq a_{n+1} \leq a_n + \beta_n, \quad n = 0, 1, \dots$$

with  $\sum_{n=0}^{\infty} \beta_n < \infty$ . It is easily seen that

$$0 \leq a_n \leq a_0 + \sum_{k=0}^{n-1} \beta_k \leq a_0 + \sum_{k=0}^{\infty} \beta_k < \infty,$$

i.e.  $\{a_n\}$  is a bounded sequence. Thus  $\limsup_{n \rightarrow \infty} a_n$  exists and is finite. Since for  $k < n$  we have  $a_n \leq a_k + \sum_{j=k}^{n-1} \beta_j$ , we may derive that  $\limsup_{n \rightarrow \infty} a_n \leq a_k + \sum_{j=k}^{\infty} \beta_j$ . This then implies that

$$\liminf_{k \rightarrow \infty} a_k \geq \limsup_{n \rightarrow \infty} a_n - \lim_{k \rightarrow \infty} \sum_{j=k}^{\infty} \beta_j = \limsup_{n \rightarrow \infty} a_n.$$

Therefore  $\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n$ , i.e.  $\lim_{n \rightarrow \infty} a_n$  exists.  $\square$

**Lemma 3.4.** *For all  $n \geq 0$  there holds*

$$\|x_n - \nabla\Theta^*(\xi_n)\|^p \leq \frac{p}{2c_0} \varepsilon_n.$$

*Proof.* Recall  $\xi_n \in \partial_{\varepsilon_n} \Theta(x_n)$ . This result follows from Lemma 2.3 immediately.  $\square$

**Lemma 3.5.** *Let all the conditions in Lemma 3.2 hold. Then there is a universal constant  $C$  such that*

$$\|F_i(x_n) - y_i\|^p \leq C \sum_{k=\min\{n, n+i-i_n\}}^{\max\{n, n+i-i_n\}} (\|r_k\|^p + \varepsilon_k)$$

for all  $n \geq 0$  and  $i = 0, \dots, N-1$ , where  $i_n = n \pmod{N}$ .

*Proof.* We consider the case  $i \geq i_n$ ; the case  $i < i_n$  can be considered similarly. Let  $d = (n - i_n)/N$ . By the triangle inequality and (3.5) we have

$$\begin{aligned} \|F_i(x_n) - y_i\| &= \|F_i(x_{Nd+i_n}) - y_i\| \\ &\leq \|F_i(x_{Nd+i}) - y_i\| + \|F_i(x_{Nd+i}) - F_i(x_{Nd+i_n})\| \\ &\leq \|r_{Nd+i}\| + \sum_{j=i_n}^{i-1} \|F_i(x_{Nd+j+1}) - F_i(x_{Nd+j})\| \\ &\leq \|r_{Nd+i}\| + \frac{B}{1-\gamma} \sum_{j=i_n}^{i-1} \|x_{Nd+j+1} - x_{Nd+j}\|. \end{aligned}$$

By virtue of the Hölder inequality and  $i - i_n \leq N-1$ , we can obtain

$$\begin{aligned} \|F_i(x_n) - y_i\|^p &\leq N^{p-1} \left( \|r_{Nd+i}\|^p + \frac{B^p}{(1-\gamma)^p} \sum_{j=i_n}^{i-1} \|x_{Nd+j+1} - x_{Nd+j}\|^p \right). \end{aligned} \quad (3.27)$$

In view of Lemma 3.4 and (2.5), we have

$$\begin{aligned} &\|x_{Nd+j+1} - x_{Nd+j}\|^p \\ &\leq 3^{p-1} (\|x_{Nd+j+1} - \nabla \Theta^*(\xi_{Nd+j+1})\|^p + \|x_{Nd+j} - \nabla \Theta^*(\xi_{Nd+j})\|^p \\ &\quad + \|\nabla \Theta^*(\xi_{Nd+j+1}) - \nabla \Theta^*(\xi_{Nd+j})\|^p) \\ &\leq \frac{p3^{p-1}}{2c_0} (\varepsilon_{Nd+j} + \varepsilon_{Nd+j+1}) + 3^{p-1} \left( \frac{\|\xi_{Nd+j+1} - \xi_{Nd+j}\|}{2c_0} \right)^{p^*}. \end{aligned}$$

By the definition of  $\xi_{Nd+j+1}$  and  $\mu_{Nd+j}$  we have

$$\begin{aligned} \|\xi_{Nd+j+1} - \xi_{Nd+j}\| &= \mu_{Nd+j} \|L_j(x_{Nd+j})^* J_s^{\mathcal{Y}_j}(r_{Nd+j})\| \\ &\leq B \mu_{Nd+j} \|r_{Nd+j}\|^{s-1} \\ &\leq B \beta_1 (\|r_{Nd+j}\|^p + \sigma \varepsilon_{Nd+j})^{1-\frac{1}{p}}. \end{aligned}$$

Thus

$$\begin{aligned} &\|x_{Nd+j+1} - x_{Nd+j}\|^p \\ &\leq \frac{p3^{p-1}}{2c_0} (\varepsilon_{Nd+j} + \varepsilon_{Nd+j+1}) + 3^{p-1} \left( \frac{B\beta_1}{2c_0} \right)^{p^*} (\|r_{Nd+j}\|^p + \sigma \varepsilon_{Nd+j}). \end{aligned}$$

Combining this with (3.27) gives

$$\begin{aligned} \|F_i(x_n) - y_i\|^p &\leq C \left( \|r_{Nd+i}\|^p + \sum_{j=i_n}^{i-1} (\|r_{Nd+j}\|^p + \varepsilon_{Nd+j+1} + \varepsilon_{Nd+j}) \right) \\ &\leq C \sum_{k=n}^{n+i-i_n} (\|r_k\|^p + \varepsilon_k), \end{aligned}$$

where  $C$  is a universal constant. This completes the proof.  $\square$

**Corollary 3.1.** *Let all the conditions in Lemma 3.2 hold. Then*

$$\lim_{n \rightarrow \infty} \|F_i(x_n) - y_i\| = 0$$

for all  $i = 0, \dots, N-1$ .

*Proof.* This result follows from Lemma 3.5 and (3.26).  $\square$

Now we are ready to prove the convergence on the sequence  $\{x_n\}$  defined by Algorithm 3.2.

**Theorem 3.6.** *Let all the conditions in Lemma 3.2 hold. Then for Algorithm 3.2 there exists a solution  $x_* \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  of (3.1) such that*

$$\lim_{n \rightarrow \infty} \|x_n - x_*\| = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} D_{\xi_n}^{\varepsilon_n} \Theta(x_*, x_n) = 0.$$

*Proof.* We first show that  $\{x_n\}$  has a convergent subsequence. To this end, we consider.

$$R_d := \sum_{n=Nd}^{Nd+N-1} (\|r_n\|^p + \varepsilon_n), \quad d = 0, 1, \dots$$

Then  $R_d > 0$  and, in view of (3.26), we have  $R_d \rightarrow 0$  as  $d \rightarrow \infty$ . We may choose a strictly increasing subsequence  $\{d_k\}$  of integers such that  $d_0 = 0$  and  $d_k$ , for each  $k \geq 1$ , is the first integer satisfying

$$d_k \geq d_{k-1} + 1 \quad \text{and} \quad R_{d_k} \leq R_{d_{k-1}}.$$

For this  $\{d_k\}$  it can be seen that

$$R_d \geq R_{d_k} \quad \forall d \leq d_k. \quad (3.28)$$

Indeed, for any  $d$  satisfying  $0 \leq d < d_k$ , we can find  $0 \leq l < k$  such that  $d_l \leq d < d_{l+1}$  and thus, by the definition of  $d_{l+1}$ , we have  $R_d \geq R_{d_l} \geq R_{d_k}$ .

With the above chosen  $\{d_k\}$ , we set  $\{n_k\} := \{Nd_k\}$  and show that  $\{x_{n_k}\}$  is convergent. To this end, we consider the  $\varepsilon$ -Bregman distance  $D_{\xi_{n_k}}^{\varepsilon_{n_k}} \Theta(x_{n_l}, x_{n_k})$  for  $k < l$ . Let  $\hat{x}$  be any solution of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ . By the definition of  $\varepsilon$ -Bregman distance, we have

$$\begin{aligned} D_{\xi_{n_k}}^{\varepsilon_{n_k}} \Theta(x_{n_l}, x_{n_k}) &= D_{\xi_{n_k}}^{\varepsilon_{n_k}} \Theta(\hat{x}, x_{n_k}) - D_{\xi_{n_l}}^{\varepsilon_{n_l}} \Theta(\hat{x}, x_{n_l}) \\ &\quad + \langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - \hat{x} \rangle + \varepsilon_{n_l}. \end{aligned} \quad (3.29)$$

We need to consider the term  $\langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - \hat{x} \rangle$  for  $k < l$ . We write

$$\begin{aligned} \langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - \hat{x} \rangle &= \langle \xi_{Nd_l} - \xi_{Nd_k}, x_{Nd_l} - \hat{x} \rangle \\ &= \sum_{d=d_k}^{d_l-1} \langle \xi_{N(d+1)} - \xi_{Nd}, x_{Nd_l} - \hat{x} \rangle. \end{aligned}$$

According to the definition of  $\xi_{n+1}$  we have

$$\begin{aligned}\langle \xi_{N(d+1)} - \xi_{Nd}, x_{Nd_l} - \hat{x} \rangle &= \sum_{n=Nd}^{Nd+N-1} \langle \xi_{n+1} - \xi_n, x_{Nd_l} - \hat{x} \rangle \\ &= \sum_{n=Nd}^{Nd+N-1} \mu_n \langle J_s^{\mathcal{Y}_{i_n}}(r_n), L_{i_n}(x_n)(x_{Nd_l} - \hat{x}) \rangle.\end{aligned}$$

By the Cauchy-Schwarz inequality, the property of  $J_s^{\mathcal{Y}_{i_n}}$  and the definition of  $\mu_n$  we obtain

$$\begin{aligned}|\langle \xi_{N(d+1)} - \xi_{Nd}, x_{Nd_l} - \hat{x} \rangle| &\leq \sum_{n=Nd}^{Nd+N-1} \mu_n \|r_n\|^{s-1} \|L_{i_n}(x_n)(x_{Nd_l} - \hat{x})\| \\ &\leq \beta_1 \sum_{n=Nd}^{Nd+N-1} (\|r_n\|^p + \sigma \varepsilon_n)^{1-\frac{1}{p}} \|L_{i_n}(x_n)(x_{Nd_l} - \hat{x})\|.\end{aligned}$$

By Assumption 3.1 (c) we have

$$\begin{aligned}\|L_{i_n}(x_n)(x_{Nd_l} - \hat{x})\| &\leq \|L_{i_n}(x_n)(x_n - \hat{x})\| + \|L_{i_n}(x_n)(x_{Nd_l} - x_n)\| \\ &\leq (1 + \gamma) (\|F_{i_n}(x_n) - F_{i_n}(\hat{x})\| + \|F_{i_n}(x_{Nd_l}) - F_{i_n}(x_n)\|) \\ &\leq (1 + \gamma) (2\|r_n\| + \|F_{i_n}(x_{Nd_l}) - y_{i_n}\|).\end{aligned}$$

Combining the above two equations and using the Hölder inequality, we can find a universal constant  $C$  such that

$$\begin{aligned}&|\langle \xi_{N(d+1)} - \xi_{Nd}, x_{Nd_l} - \hat{x} \rangle| \\ &\leq CR_d + C \sum_{n=Nd}^{Nd+N-1} (\|r_n\|^p + \varepsilon_n)^{1-\frac{1}{p}} \|F_{i_n}(x_{Nd_l}) - y_{i_n}\| \\ &\leq CR_d + CR_d^{1/p^*} \left( \sum_{n=Nd}^{Nd+N-1} \|F_{i_n}(x_{Nd_l}) - y_{i_n}\|^p \right)^{1/p}.\end{aligned}$$

By using Lemma 3.5 we have

$$\begin{aligned}\sum_{n=Nd}^{Nd+N-1} \|F_{i_n}(x_{Nd_l}) - y_{i_n}\|^p &\leq C \sum_{n=Nd}^{Nd+N-1} \sum_{k=Nd_l}^{Nd_l+i_n} (\|r_k\|^p + \varepsilon_k) \\ &\leq CN \sum_{k=Nd_l}^{Nd_l+N-1} (\|r_k\|^p + \varepsilon_k) \\ &= CN R_{d_l}.\end{aligned}$$

Consequently, it follows from (3.28) that

$$|\langle \xi_{N(d+1)} - \xi_{Nd}, x_{Nd_l} - \hat{x} \rangle| \leq CR_d + CR_d^{1/p^*} R_{d_l}^{1/p} \leq CR_d.$$

Therefore

$$|\langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - \hat{x} \rangle| \leq C \sum_{d=d_k}^{d_l-1} R_d = C \sum_{n=n_k}^{n_l-1} (\|r_n\|^p + \varepsilon_n).$$

This together with (3.26) implies that

$$\lim_{k \rightarrow \infty} \sup_{l \geq k} |\langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - \hat{x} \rangle| = 0. \quad (3.30)$$

Thus, it follows from (3.29), Lemma 3.3 and  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  that

$$D_{\xi_{n_k}}^{\varepsilon_{n_k}} \Theta(x_{n_l}, x_{n_k}) \rightarrow 0 \quad \text{as } k, l \rightarrow \infty.$$

In view of Lemma 2.1 and  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ , we can conclude

$$\|x_{n_l} - x_{n_k}\| \rightarrow 0 \quad \text{as } k, l \rightarrow \infty,$$

i.e.  $\{x_{n_k}\}$  is a Cauchy sequence in  $\mathcal{X}$ . Thus  $x_{n_k} \rightarrow x_*$  as  $k \rightarrow \infty$  for some  $x_* \in \mathcal{X}$ . By using Corollary 3.1 and the continuity of  $F_i$  we have  $F_i(x_*) = y_i$  for all  $i = 0, \dots, N-1$ .

We next show that  $x_* \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ . Since  $\{x_{n_k}\} \subset B_{2\rho}(x_0)$ , we must have  $x_* \in B_{2\rho}(x_0)$ . By using  $\xi_{n_k} \in \partial_{\varepsilon_{n_k}} \Theta(x_{n_k})$  we can obtain

$$\Theta(x_{n_k}) \leq \Theta(\hat{x}) + \langle \xi_{n_k}, x_{n_k} - \hat{x} \rangle + \varepsilon_{n_k}. \quad (3.31)$$

In view of (3.30) and  $x_{n_k} \rightarrow x_*$ , there is a constant  $C_0$  such that

$$|\langle \xi_{n_k} - \xi_{n_0}, x_{n_k} - \hat{x} \rangle| \leq C_0 \quad \text{and} \quad |\langle \xi_{n_0}, x_{n_k} - \hat{x} \rangle| \leq C_0, \quad \forall k.$$

Thus  $|\langle \xi_{n_k}, x_{n_k} - \hat{x} \rangle| \leq 2C_0$  for all  $k$ . By using the lower semi-continuity of  $\Theta$  and  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  we obtain from (3.31) that

$$\Theta(x_*) \leq \liminf_{k \rightarrow \infty} \Theta(x_{n_k}) \leq \Theta(\hat{x}) + 2C_0 < \infty.$$

This implies that  $x_* \in \mathcal{D}(\Theta)$ . Therefore  $x_*$  is a solution of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ .

Finally we show that the whole sequence  $\{x_n\}$  converges to  $x_*$ . Let

$$\eta_0 := \lim_{n \rightarrow \infty} D_{\xi_n}^{\varepsilon_n} \Theta(x_*, x_n)$$

whose existence is guaranteed by Lemma 3.3. By the non-negativity of  $\varepsilon$ -Bregman distance, we have  $\eta_0 \geq 0$ . In (3.29) we set  $\hat{x} = x_*$  and take  $l \rightarrow \infty$  to derive that

$$D_{\xi_{n_k}}^{\varepsilon_{n_k}} \Theta(x_*, x_{n_k}) \leq D_{\xi_{n_k}}^{\varepsilon_{n_k}} \Theta(x_*, x_{n_k}) - \eta_0 + \sup_{l \geq k} |\langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - x_* \rangle|.$$

This implies that

$$\eta_0 \leq \sup_{l \geq k} |\langle \xi_{n_l} - \xi_{n_k}, x_{n_l} - x_* \rangle|$$

for all  $k$ . In view of (3.30), by taking  $k \rightarrow \infty$  we obtain  $\eta_0 \leq 0$ . Therefore  $\eta_0 = 0$ , that is,  $\lim_{n \rightarrow \infty} D_{\xi_n}^{\varepsilon_n} \Theta(x_*, x_n) = 0$ . By using Lemma 2.1 and  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$  we can obtain  $\lim_{n \rightarrow \infty} \|x_n - x_*\| = 0$ .  $\square$

**3.3. Regularization property.** We return to Algorithm 3.1 and prove its regularization property. We need the following stability result.

**Lemma 3.7.** *Let all the conditions in Lemma 3.1 hold. Then for all  $n \geq 0$  there hold*

$$\xi_n^\delta \rightarrow \xi_n \quad \text{and} \quad x_n^\delta \rightarrow x_n \quad \text{as } \delta \rightarrow 0.$$

*Proof.* The result is trivial for  $n = 0$ . We next assume that the result is true for some  $n \geq 0$  and show that  $\xi_{n+1}^\delta \rightarrow \xi_{n+1}$  and  $x_{n+1}^\delta \rightarrow x_{n+1}$  as  $\delta \rightarrow 0$ . We consider two cases.

*Case 1:*  $F_{i_n}(x_n) = y_{i_n}$ . In this case we have  $\xi_{n+1} = \xi_n$ . Therefore

$$\xi_{n+1}^\delta - \xi_{n+1} = \xi_n^\delta - \xi_n - \mu_n^\delta L_{i_n}(x_n^\delta)^* J_s^{\mathcal{Y}_{i_n}}(r_n^\delta).$$

Since  $\mu_n^\delta \leq \beta_1(\|r_n^\delta\|^p + \sigma\varepsilon_n)^{1-\frac{s}{p}}$ , we may use the property of  $J_s^{\mathcal{Y}_{i_n}}$  and (3.4) to obtain

$$\|\xi_{n+1}^\delta - \xi_{n+1}\| \leq \|\xi_n^\delta - \xi_n\| + \beta_1 B (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{1-\frac{s}{p}} \|r_n^\delta\|^{s-1}.$$

By the induction hypothesis and the continuity of  $F_{i_n}$ , we then have  $\xi_{n+1}^\delta \rightarrow \xi_{n+1}$  as  $\delta \rightarrow 0$ . By the definition of  $x_{n+1}^\delta$  and the continuity of  $S_\varepsilon$  we have

$$x_{n+1}^\delta = S_{\varepsilon_{n+1}}(\xi_{n+1}^\delta) \rightarrow S_{\varepsilon_{n+1}}(\xi_{n+1}) = x_{n+1}$$

as  $\delta \rightarrow 0$ .

*Case 2:*  $F_{i_n}(x_n) \neq y_{i_n}$ . Since  $\varepsilon_n > 0$ , we have  $\|r_n^\delta\|^p + \sigma\varepsilon_n > (\tau\delta)^p$  for small  $\delta > 0$ . Thus

$$\mu_n = \tilde{\mu}_n (\|r_n\|^p + \sigma\varepsilon_n)^{1-\frac{s}{p}} \quad \text{and} \quad \mu_n^\delta = \tilde{\mu}_n^\delta (\|r_n^\delta\|^p + \sigma\varepsilon_n)^{1-\frac{s}{p}}.$$

We claim that  $\mu_n^\delta \rightarrow \mu_n$  as  $\delta \rightarrow 0$ . In fact, if  $L_{i_n}(x_n)^* J_s^{\mathcal{Y}_{i_n}}(r_n) = 0$ , then, by definition, we must have  $\tilde{\mu}_n = \beta_1$  and  $\tilde{\mu}_n^\delta = \beta_1$  for small  $\delta > 0$ . If  $L_{i_n}(x_n)^* J_s^{\mathcal{Y}_{i_n}}(r_n) \neq 0$ , then, by the definition of  $\tilde{\mu}_n^\delta$  and  $\tilde{\mu}_n$  and the induction hypothesis, we can conclude that  $\tilde{\mu}_n^\delta \rightarrow \tilde{\mu}_n$  as  $\delta \rightarrow 0$ . In any case, we always have  $\tilde{\mu}_n^\delta \rightarrow \tilde{\mu}_n$  as  $\delta \rightarrow 0$ . Consequently, by the induction hypotheses and the continuity of  $F_{i_n}$ ,  $L_{i_n}$  and  $J_s^{\mathcal{Y}_{i_n}}$ , we have  $\mu_n^\delta \rightarrow \mu_n$  and therefore  $\xi_{n+1}^\delta \rightarrow \xi_{n+1}$  as  $\delta \rightarrow 0$ . By invoking again the continuity of  $S_\varepsilon$ , we obtain  $x_{n+1}^\delta \rightarrow x_{n+1}$  as  $\delta \rightarrow 0$ .  $\square$

**Theorem 3.8.** *Assume that all the conditions in Lemma 3.1 hold. Then for Algorithm 3.1 there hold*

$$\lim_{\delta \rightarrow 0} \|x_{n_\delta}^\delta - x_*\| = 0 \quad \text{and} \quad \lim_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta}^{\varepsilon_{n_\delta}} \Theta(x_*, x_{n_\delta}^\delta) = 0,$$

where  $x_*$  is a solution of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ .

*Proof.* By the definition of  $n_\delta$  we have  $\sigma\varepsilon_{n_\delta} \leq (\tau\delta)^p$ . Since  $\varepsilon_n > 0$  for all  $n$ , we must have  $n_\delta \rightarrow \infty$  as  $\delta \rightarrow 0$ . Let  $x_*$  be the solution of (3.1) in  $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$  determined in Theorem 3.6. In view of Lemma 2.1 and  $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ , it suffices to show that

$$\lim_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta}^{\varepsilon_{n_\delta}} \Theta(x_*, x_{n_\delta}^\delta) = 0. \quad (3.32)$$

Let  $n$  be an arbitrary but fixed integer. By virtue of Lemma 3.1 we have

$$D_{\xi_{n_\delta}^\delta}^{\varepsilon_{n_\delta}} \Theta(x_*, x_{n_\delta}^\delta) \leq D_{\xi_n^\delta}^{\varepsilon_n} \Theta(x_*, x_n^\delta) + 3 \sum_{k=n}^{n_\delta} \varepsilon_k.$$

Consequently

$$\limsup_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta}^{\varepsilon_{n_\delta}} \Theta(x_*, x_{n_\delta}^\delta) \leq \limsup_{\delta \rightarrow 0} D_{\xi_n^\delta}^{\varepsilon_n} \Theta(x_*, x_n^\delta) + 3 \sum_{k=n}^{\infty} \varepsilon_k.$$

By making use of Lemma 3.7 and the lower semi-continuity of  $\Theta$  we have

$$\limsup_{\delta \rightarrow 0} D_{\xi_n^\delta}^{\varepsilon_n} \Theta(x_*, x_n^\delta) \leq D_{\xi_n}^{\varepsilon_n} \Theta(x_*, x_n).$$

Therefore

$$\limsup_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta}^{\varepsilon_{n_\delta}} \Theta(x_*, x_{n_\delta}^\delta) \leq D_{\xi_n}^{\varepsilon_n} \Theta(x_*, x_n) + 3 \sum_{k=n}^{\infty} \varepsilon_k.$$

Since  $n$  can be arbitrarily large, by taking  $n \rightarrow \infty$  we can derive from Theorem 3.6 and the condition  $\sum_{k=0}^{\infty} \varepsilon_k < \infty$  that

$$\limsup_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta}^{\varepsilon_{n_\delta}} \Theta(x_*, x_{n_\delta}^\delta) \leq 0.$$



This completes the proof.  $\square$

**3.4. Acceleration.** It is well-known that Landweber iteration is a slowly convergent method ([10]). In order to make it applicable in practical applications, acceleration strategies should be incorporated into the method. When an inverse problem is formulated in Hilbert spaces, a family of accelerated Landweber iterations, including the famous  $\nu$ -method of Brakhage [2], have been proposed in [3] using the orthogonal polynomials and the spectral theory of self-adjoint operators. The acceleration strategy using orthogonal polynomials is no longer available when an inverse problems is considered in Banach spaces using general convex penalty functions. Instead the sequential subspace optimization strategy has been employed in [16, 32] to accelerate the method.

In recent years Nesterov's acceleration strategy [29] has received tremendous consideration in optimization community. Consider the unconstrained optimization problem

$$\min_{x \in \mathcal{X}} \varphi(x) \quad (3.33)$$

in a Hilbert space  $\mathcal{X}$  with a continuous differentiable function  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ . Nesterov's strategy speeds up the gradient descent method by using a proper extrapolation point at each iteration step. It takes the form

$$\begin{cases} \hat{x}_n = x_n + \frac{n}{n+\alpha}(x_n - x_{n-1}), \\ x_{n+1} = \hat{x}_n - t_n \nabla \varphi(\hat{x}_n) \end{cases} \quad (3.34)$$

with suitable step sizes  $t_n > 0$ , where  $\alpha \geq 3$  is a fixed number. This strategy has been extended in various context, even for nonsmooth optimization problems, see [2, 3, 8].

Due to the simplicity of Nesterov's strategy, it is natural to consider its use in accelerating our Landweber-Kaczmarz method. We propose the following accelerated version in which we drop the superscript  $\delta$  for all the iterates for simplicity.

**Algorithm 3.3.** Let  $\beta_0 > 0$ ,  $\beta_1 > 0$ ,  $\sigma > 0$  and  $\tau > 1$  be suitably chosen numbers, let  $\{\varepsilon_n\}$  be a summable sequence of positive numbers, and let  $\alpha \geq 3$  be a fixed number.

- (i) Pick  $x_0 \in \mathcal{X}$  and  $\xi_0 \in \mathcal{X}^*$  such that  $\xi_0 \in \partial\Theta(x_0)$ .
- (ii) Let  $\xi_{-1} = \xi_0$ ,  $x_{-1} = x_0$ ,  $q_{-1} = 0$  and  $\hat{x}_0 = x_0$ . For  $n \geq 0$  we define  $\hat{r}_n = F_{i_n}(\hat{x}_n) - y_{i_n}^\delta$  and

$$q_n = \begin{cases} q_{n-1} + 1 & \text{if } \|\hat{r}_n\|^p + \sigma\varepsilon_n \leq (\tau\delta)^p, \\ 0 & \text{otherwise.} \end{cases}$$

We then update

$$\begin{cases} \hat{\xi}_n = \xi_n + \frac{n}{n+\alpha}(\xi_n - \xi_{n-1}), \\ \hat{x}_n = x_n + \frac{n}{n+\alpha}(x_n - x_{n-1}), \\ \xi_{n+1} = \hat{\xi}_n - \mu_n L_{i_n}(\hat{x}_n)^* J_s^{\mathcal{Y}_{i_n}}(\hat{r}_n), \\ x_{n+1} = S_{\varepsilon_{n+1}}(\xi_{n+1}), \end{cases} \quad (3.35)$$

where

$$\mu_n = \begin{cases} \tilde{\mu}_n (\|\hat{r}_n\|^p + \sigma\varepsilon_n)^{1-\frac{\sigma}{p}} & \text{if } \|\hat{r}_n\|^p + \sigma\varepsilon_n > (\tau\delta)^p, \\ 0 & \text{otherwise} \end{cases}$$

with

$$\tilde{\mu}_n = \min \left\{ \frac{\beta_0 \|\hat{r}_n\|^{p(s-1)}}{\|L_{i_n}(\hat{x}_n)^* J_r^{\mathcal{Y}_{i_n}}(\hat{r}_n)\|^p}, \beta_1 \right\};$$

(iii) Let  $n_\delta$  be the first integer such that  $q_{n_\delta} = N$  and use  $x_{n_\delta}$  as an approximate solution.

Let us give a brief remark on Algorithm 3.3 when  $N = 1$ . Note that, when both  $\mathcal{X}$  and  $\mathcal{Y}_0$  are Hilbert spaces,  $F_0 : \mathcal{X} \rightarrow \mathcal{Y}$  is a bounded linear operator and  $\Theta(x) = \frac{1}{2}\|x\|^2$ , by taking  $J_s^{\mathcal{Y}_0} = J_2^{\mathcal{Y}_0} \equiv \text{id}$  and let  $S_\varepsilon$  be the exact solver, i.e.  $S_\varepsilon(\xi) = \xi$  for any  $\xi \in \mathcal{X}$  and  $\varepsilon > 0$ , then (3.35) becomes

$$\begin{cases} \hat{x}_n = x_n + \frac{n}{n+\alpha}(x_n - x_{n-1}), \\ x_{n+1} = \hat{x}_n - \mu_n F_0^* (F_0 \hat{x}_n - y_0^\delta). \end{cases} \quad (3.36)$$

This is exactly the formula (3.34) applied to (3.33) with  $\varphi(x) = \frac{1}{2}\|F_0 x - y_0^\delta\|^2$ . Therefore, it is reasonable to expect that Algorithm 3.3 can converge faster than Algorithm 3.1. Currently there is no available theory regarding the convergence of Algorithm 3.3, however, in the next section we will use numerical examples to demonstrate its acceleration effect.

**3.5. Construction of inexact inner solvers.** In this subsection we will discuss how to find an inexact solver specified in Assumption 3.2 for solving (3.6) for each  $\varepsilon > 0$ . For those  $\Theta$  such that the exact solution of (3.6) can be determined explicitly, we can simply take each  $S_\varepsilon$  to be the exact solver. We therefore consider only those  $\Theta$  for which (3.6) does not have an explicit exact solution. We will focus on the total variation like functions which have significant importance in image reconstruction.

Due to the numerical implementation, we will give the exposition in a discrete setting. Let  $\mathcal{X} = \mathbb{R}^{I \times J}$  and for each  $z \in \mathcal{X}$  let  $\|z\|_F$  denotes the Fröbenius norm of  $z$ . Let  $D : \mathcal{X} \rightarrow \mathcal{X} \times \mathcal{X}$  be the discrete gradient operator given by  $Dz = (D_1 z, D_2 z)$  for  $z = (z_{i,j}) \in \mathcal{X}$  with

$$(D_1 z)_{i,j} = \begin{cases} z_{i+1,j} - z_{i,j} & \text{if } i = 1, \dots, I-1; j = 1, \dots, J, \\ z_{1,j} - z_{I,j} & \text{if } i = I; j = 1, \dots, J \end{cases}$$

and

$$(D_2 z)_{i,j} = \begin{cases} z_{i,j+1} - z_{i,j} & \text{if } i = 1, \dots, I; j = 1, \dots, J-1, \\ z_{i,1} - z_{i,J} & \text{if } i = 1, \dots, I; j = J. \end{cases}$$

We consider the function

$$\Theta(z) = \frac{1}{2\mu} \|z\|_F^2 + h(Dz) + \iota_{\mathcal{C}}(z), \quad z \in \mathcal{X}, \quad (3.37)$$

where  $\mu > 0$  is a constant,  $\mathcal{C} \subset \mathcal{X}$  is a closed convex set representing the constraints on  $z$ ,  $\iota_{\mathcal{C}}$  is the indicator function of  $\mathcal{C}$ , i.e.

$$\iota_{\mathcal{C}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{C}, \\ +\infty & \text{otherwise,} \end{cases}$$

and  $h$  is a function defined on  $\mathcal{X} \times \mathcal{X}$  given by

$$h(u, v) = \sum_{i=1}^I \sum_{j=1}^J \sqrt{u_{i,j}^2 + v_{i,j}^2}, \quad (u, v) \in \mathcal{X} \times \mathcal{X}.$$

Note that  $|z|_{TV} := h(Dz)$  represents a discrete total variation of  $z$ . For this  $\Theta$  the minimization problem (3.6) becomes

$$x = \arg \min_{z \in \mathcal{X}} \left\{ \Psi_P(z) := \frac{1}{2\mu} \|z - \mu\xi\|_F^2 + h(Dz) + \iota_{\mathcal{C}}(z) \right\}, \quad (3.38)$$

where  $\Psi_P(z)$  is called a primal function. This is a total variation denoising problem [30] for which many algorithms have been developed to solve it approximately. We will use the primal-dual hybrid gradient (PDHG) method introduced in [36] which is a special case of the Uzawa algorithm [1]. To formulate the method, we use the Legendre-Fenchel conjugate  $h^*$  of  $h$  to rewrite (3.38) as

$$x = \arg \min_{z \in \mathcal{X}} \max_{\lambda \in \mathcal{X} \times \mathcal{X}} \left\{ \Phi(z, \lambda) := \frac{1}{2\mu} \|z - \mu\xi\|_F^2 + \langle \lambda, Dz \rangle_F - h^*(\lambda) + \iota_{\mathcal{C}}(z) \right\}.$$

Then the PDHG method takes the form

$$\begin{aligned} \lambda_{k+1} &= \arg \max_{\lambda \in \mathcal{X} \times \mathcal{X}} \left\{ \Phi(z_k, \lambda) - \frac{1}{2\mu\tau_k} \|\lambda - \lambda_k\|_F^2 \right\}, \\ z_{k+1} &= \arg \min_{z \in \mathcal{X}} \left\{ \Phi(z, \lambda_{k+1}) + \frac{1 - \theta_k}{2\mu\theta_k} \|z - z_k\|_F^2 \right\} \end{aligned}$$

with suitably chosen step sizes  $\tau_k > 0$  and  $0 < \theta_k < 1$ . Direct manipulation shows that

$$\begin{aligned} \lambda_{k+1} &= \text{prox}_{\mu\tau_k h^*}(\lambda_k + \mu\tau_k Dz_k), \\ z_{k+1} &= \Pi_{\mathcal{C}}((1 - \theta_k)z_k + \mu\theta_k(\xi - D^T \lambda_{k+1})), \end{aligned}$$

where  $\Pi_{\mathcal{C}}$  denotes the projection onto  $\mathcal{C}$  and  $\text{prox}_{th^*}$  for any  $t > 0$  denotes the proximal mapping of  $h^*$  defined by

$$\text{prox}_{th^*}(\bar{\lambda}) := \arg \min_{\lambda \in \mathcal{X} \times \mathcal{X}} \left\{ h^*(\lambda) + \frac{1}{2t} \|\lambda - \bar{\lambda}\|_F^2 \right\}, \quad \bar{\lambda} \in \mathcal{X} \times \mathcal{X}.$$

For our  $h$ , it is easily seen that  $h^*(\lambda) = \iota_{\mathcal{Z}}(\lambda)$ , where

$$\mathcal{Z} := \{(u, v) \in \mathcal{X} \times \mathcal{X} : u_{i,j}^2 + v_{i,j}^2 \leq 1 \text{ for } i = 1, \dots, I; j = 1, \dots, J\}.$$

Thus  $\text{prox}_{th^*}(\lambda) = \Pi_{\mathcal{Z}}(\lambda)$ , where, for any  $\lambda := (u, v) \in \mathcal{X} \times \mathcal{X}$ ,  $\Pi_{\mathcal{Z}}(\lambda) = (y, z)$  with

$$y_{i,j} = \frac{u_{i,j}}{\max\{1, \sqrt{u_{i,j}^2 + v_{i,j}^2}\}}, \quad z_{i,j} = \frac{v_{i,j}}{\max\{1, \sqrt{u_{i,j}^2 + v_{i,j}^2}\}}.$$

To achieve a fast convergence, it was suggested in [36] to choose the step sizes as

$$\tau_k = 0.2 + 0.08k, \quad \theta_k = \left(0.5 - \frac{5}{15 + k}\right) / \tau_k.$$

The convergence of the PDHG method, under such a choice of the step sizes, was confirmed in [4].

In order to terminate the PDHG method, we use the relative duality gap. The dual function is given by

$$\begin{aligned} \Psi_D(\lambda) &= \min_{z \in \mathcal{X}} \Phi(z, \lambda) \\ &= \begin{cases} \frac{1}{2\mu} \|(I - \Pi_{\mathcal{C}})(\mu\xi - \mu D^T \lambda)\|_F^2 + \frac{\mu}{2} \|\xi - D^T \lambda\|_F^2, & \lambda \in \mathcal{Z}, \\ -\infty, & \lambda \notin \mathcal{Z}. \end{cases} \end{aligned}$$

Given a primal feasible point  $z \in \mathcal{X}$  and a dual feasible point  $\lambda \in \mathcal{X} \times \mathcal{X}$ , we define the relative duality gap

$$G_{rel}(z, \lambda) = \frac{\Psi_P(z) - \Psi_D(\lambda)}{|\Psi_P(z)| + |\Psi_D(\lambda)|}.$$

If there exists a feasible  $(x, \lambda)$  such that  $G_{rel}(x, \lambda) \leq \eta$  for some  $0 < \eta < 1$ , then, by using the fact that  $\Psi_D(\lambda) \leq \min_{z \in \mathcal{X}} \Psi_P(z)$ , we can conclude that  $x$  satisfies (3.7) with

$$\varepsilon = \frac{2\eta}{1 - \eta} \min_{z \in \mathcal{X}} \Psi_P(z).$$

#### 4. Numerical simulations

In this section we provide numerical simulations to test our theoretical result on Algorithm 3.1 using inexact inner solvers and to illustrate the acceleration effect embraced in Algorithm 3.3. Our simulations were done by using MATLAB R2012a on a Lenovo laptop with Intel(R) Core(TM) i5 CPU 2.30GHz and 6GB memory.

**Example 4.1.** We first consider the application of our algorithms in computed tomography which consists in determining the density of cross sections of human body by measuring the attenuation of X-rays as they propagate through the biological tissues [28]. Mathematically, it requires to determine a function  $f$  supported on a bounded domain from its Radon transform

$$\begin{aligned} \mathcal{R}f(\theta, \rho) &:= \int_{x \cos \theta + y \sin \theta = \rho} f(x, y) ds \\ &= \int_{-\infty}^{\infty} f(\rho \cos \theta - t \sin \theta, \rho \sin \theta + t \cos \theta) dt, \end{aligned}$$

where  $\theta \in [0, 2\pi)$  and  $\rho \in \mathbb{R}$ . The most prominent method in computed tomography is the filtered backprojection (FBP) algorithm which is based on the explicit inversion formula and therefore is fast and inexpensive. However, the FBP algorithm is not robust with respect to noise, its accuracy requires patients to be over-exposed to X-rays, and it is difficult to incorporate a prior information into the algorithm.

In order to apply our algorithm to solve CT problems, we need a discrete model. We assume that the image is supported on a square domain in  $\mathbb{R}^2$  which is divided into  $q \times q$  pixels numbered from 1 to  $Q = q^2$ . Each pixel  $i$  is assigned a constant value  $f_i$  such that the vector  $f = (f_1, \dots, f_Q)^T$  is a discrete version of the sought function. Assume that there are  $M$  X-rays passing through the image. Let  $a_{ij}$  denote the length of the intersection of the  $i$ -th ray with the  $j$ -th pixel. Let  $g_i$  be the measurement of the attenuation for ray  $i$  and let  $g = (g_1, \dots, g_M)^T$ . Then

$$g = Af$$

Note that  $A$  is a sparse matrix of size  $M \times Q$ . We will apply our algorithms to solve this linear equation.

The formation of the matrix  $A$  depends on the scan geometry. In the following numerical simulations we consider only test problems that model the standard 2D parallel-beam tomography; other scan geometries can be done similarly. For the simulations, the true image  $f_*$  is taken to be the Shepp-Logan phantom shown in Figure 1(a) discretized on a  $256 \times 256$  pixel grid with its pixel values varying in the interval  $[0, 1]$ . This phantom is widely used in evaluating tomographic reconstruction algorithms. We consider a full angle problem using 45 projection angles evenly

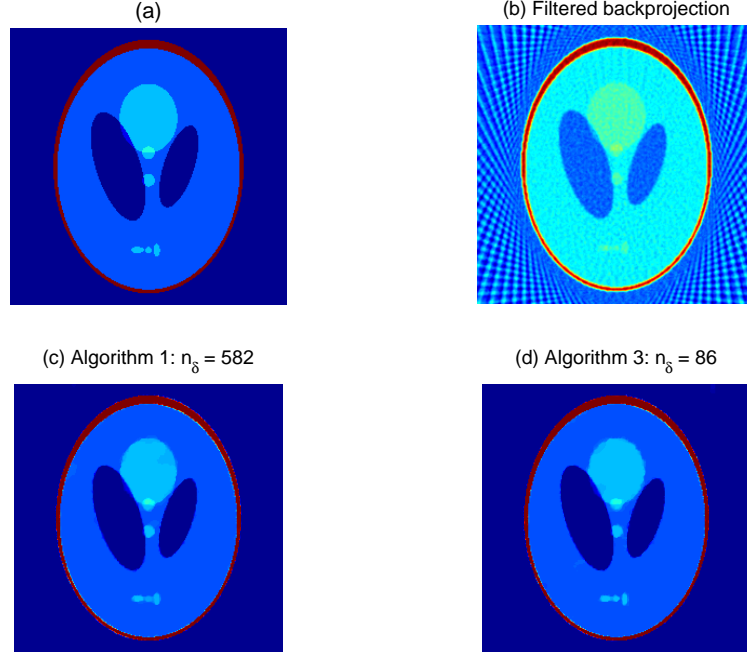


FIGURE 1. Reconstruction results by Algorithm 3.1, Algorithm 3.3 and the filtered backprojection for the computed tomography in Example 4.1.

distributed between 1 and 180 degrees, with 367 lines per projection. The function `parallel_tomo` in the MATLAB package AIR TOOLS [15] is used to generate the sparse matrix  $A$ , which has the dimension size  $M = 16515$  and  $Q = 65536$ .

Let  $g = Af_*$ . We add Gaussian noise to  $g$  to generate a noisy data  $g^\delta$  with relative noise level  $\delta_{rel} := \|g^\delta - g\|_F / \|g\|_F = 0.01$  so that the noise level is  $\delta = \delta_{rel} \|g\|_F$ . We then use  $g^\delta$  to reconstruct  $f_*$  via Algorithm 3.1 and Algorithm 3.3 with  $N = 1$ . We take

$$\Theta(f) = \frac{1}{2\mu} \|f\|_F^2 + |f|_{TV} + \iota_{\mathcal{C}}(f)$$

with  $\mu = 1$  and  $\mathcal{C} = \{f : f \geq 0\}$  and use the initial guess  $f_0 = \xi_0 = 0$  and the parameters  $\beta_0 = 0.1/\mu$ ,  $\beta_1 = 10$ ,  $\sigma = 0.001$  and  $\tau = 1.01$ ; we also take  $\alpha = 5$  when using Algorithm 3.3. The minimization problems associated with  $\Theta$  are solved by the PDHG method which is terminated when the relative duality gap is  $\leq (n+1)^{-2.2}$  at the  $n$ -th iteration. In Figure 1 (c) and (d) we present the reconstruction results by Algorithm 3.1 and Algorithm 3.3 respectively. The both algorithms give satisfactory results. Algorithm 3.1 terminates after 582 iterations and takes 386 seconds; while Algorithm 3.3 terminates after 86 iterations and takes 26.6 seconds. This clearly shows that Algorithm 3.3 is much faster than Algorithm 3.1. As comparison, in Figure 1 (b) we include the reconstruction result by the FBP algorithm which is much faster but the result is much worse.

To further illustrate the fast convergence property of Algorithm 3.3, we redo the numerical simulations under the same situation but with exact data. We run

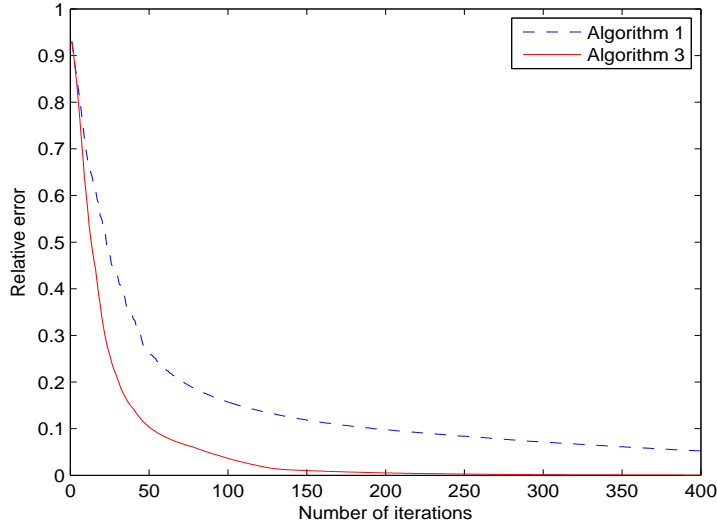


FIGURE 2. The plot of relative error of the solution versus iteration number by Algorithm 3.1 and Algorithm 3.3 for Example 4.1

400 iterations for both Algorithm 3.1 and Algorithm 3.3. The relation between the relative errors  $\|f_n - f_*\|_F / \|f_*\|_F$  and the iteration numbers is plotted in Figure 2 which clearly shows that Algorithm 3.3 has the acceleration effect.

**Example 4.2.** We next consider the identification of the parameter  $c$  in the boundary value problem

$$\begin{cases} -\Delta u + cu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega \end{cases} \quad (4.1)$$

from an  $L^2(\Omega)$ -measurement of the state  $u$ , where  $\Omega \subset \mathbb{R}^2$  is a bounded domain with Lipschitz boundary,  $f \in L^2(\Omega)$  and  $g \in H^{3/2}(\Omega)$ . This is a benchmark example of nonlinear inverse problems. We assume that the exact solution  $c^\dagger$  is in  $L^2(\Omega)$ . This problem reduces to solving  $F(c) = u$ , if we define the nonlinear operator  $F : L^2(\Omega) \rightarrow L^2(\Omega)$  by  $F(c) := u(c)$ , where  $u(c) \in H^2(\Omega) \subset L^2(\Omega)$  is the unique solution of (4.1). This operator  $F$  is well defined on

$$\mathcal{D}(F) := \{c \in L^2(\Omega) : \|c - \hat{c}\|_{L^2(\Omega)} \leq \gamma_0 \text{ for some } \hat{c} \geq 0, \text{ a.e.}\}$$

for some positive constant  $\gamma_0 > 0$ . It is known that  $F$  is Fréchet differentiable; the Fréchet derivative of  $F$  and its adjoint are given by

$$F'(c)h = -A(c)^{-1}(hF(c)) \quad \text{and} \quad F'(c)^*w = -u(c)A(c)^{-1}w \quad (4.2)$$

for  $h, w \in L^2(\Omega)$ , where  $A(c) : H^2(\Omega) \cap H_0^1(\Omega) \rightarrow L^2(\Omega)$  is defined by  $A(c)u = -\Delta u + cu$  which is an isomorphism uniformly in ball  $B_{2\rho}(c^\dagger)$  for small  $\rho > 0$ . Moreover, Assumption 3.1 (c) holds for small  $\rho > 0$  (see [10]).

We will present our numerical results on  $\Omega = [0, 1] \times [0, 1]$  with  $g \equiv 1$  on  $\partial\Omega$  and

$$f(x, y) = 200e^{-10(x-0.5)^2 - 10(y-0.5)^2} \quad \text{in } \Omega.$$

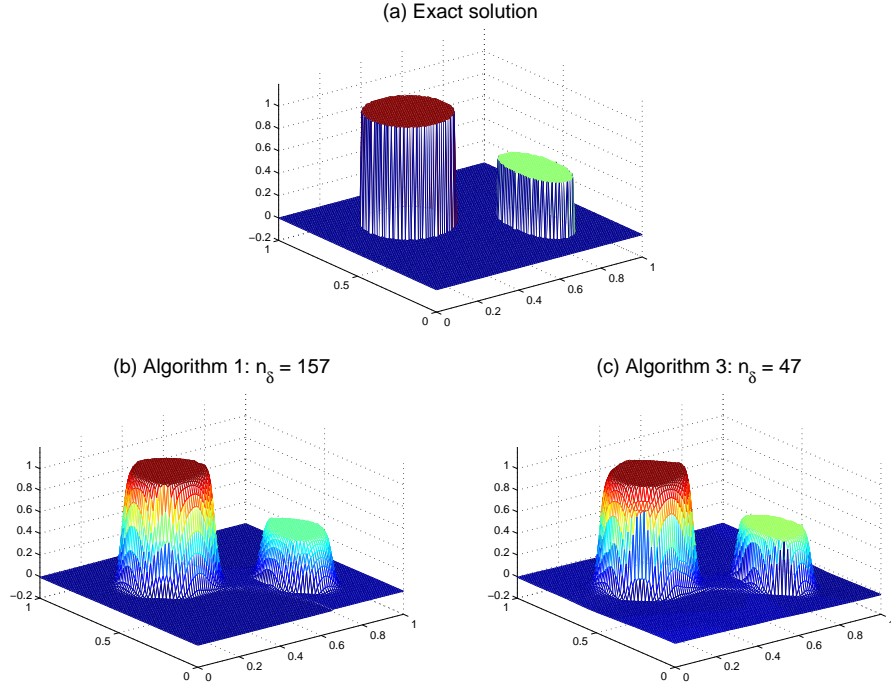


FIGURE 3. Reconstruction results by Algorithm 3.1 and Algorithm 3.3 for the parameter identification problem in Example 4.2.

We assume the sought solution  $c_*$  is a piecewise constant function as shown in Figure 3 (a) and reconstruct it by using a noisy data  $u^\delta$  with relative noise level  $\delta_{rel} := \|u^\delta - u\|_{L^2} / \|u\|_{L^2} = 0.46 \times 10^{-3}$ . When applying Algorithm 3.1 and Algorithm 3.3 with  $N = 1$ , we take  $\mathcal{X} = \mathcal{Y} = L^2(\Omega)$  and

$$\Theta(c) = \frac{1}{2\mu} \|c\|_{L^2(\Omega)}^2 + \int_{\Omega} |Dc|$$

with  $\mu = 20$ , where

$$\int_{\Omega} |Dc| = \sup \left\{ \int_{\Omega} c \operatorname{div} \varphi dx : \varphi \in C_0^1(\Omega, \mathbb{R}^2) \text{ and } \|\varphi\|_{L^\infty} \leq 1 \right\}$$

denotes the total variation of  $c$ . We use the initial guess  $c_0 = \xi_0 = 0$  and the parameters  $\beta_0 = 0.01/\mu$ ,  $\beta_1 = 2 \times 10^4$ ,  $\sigma = 0.001$  and  $\tau = 1.02$ ; we also take  $\alpha = 5$  when using Algorithm 3.3. In order to carry out the computation, we need to discretize the problem. We divide  $\Omega$  into  $100 \times 100$  small squares of equal size and solve all partial differential equations involved approximately by a finite difference method. We also discretize  $\Theta(c)$  so that  $\|c\|_{L^2}$  is replaced by the Fröbenius norm of arrays and  $\int_{\Omega} |Dc|$  is replaced by the discrete total variation given in Section 3.5. The corresponding minimization problems associated with the discrete  $\Theta$  are solved by the PDHG method which is terminated as long as the relative duality gap is  $\leq (n+1)^{-1.5}$  at the  $n$ -th iteration. In Figure 3 (b) and (c) we report the computational results using Algorithm 3.1 and Algorithm 3.3. The both algorithms give satisfactory reconstruction results. Moreover, Algorithm 3.3 requires significantly

less number of iterations than Algorithm 3.1 which demonstrates that Algorithm 3.3 has the acceleration effect.

To further illustrate the fast convergence property of Algorithm 3.3, we redo the above computation using exact data. We perform 100 iterations for both Algorithm 3.1 and Algorithm 3.3. The curve of the relative error  $\|c_n - c_*\|_{L^2}/\|c_*\|_{L^2}$  versus the iteration number  $n$  is plot in Figure 4 which clearly indicates that Algorithm 3.3 converges faster than Algorithm 3.1.

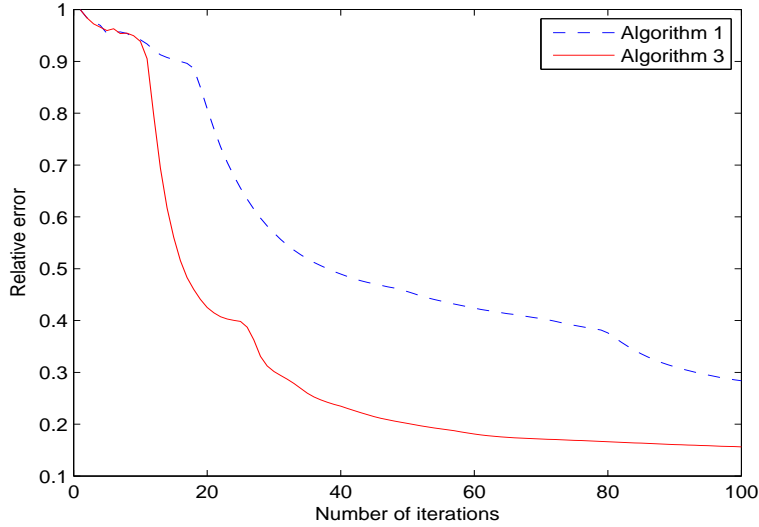


FIGURE 4. The plot of relative error of the solution versus iteration number by Algorithm 3.1 and Algorithm 3.3 for Example 4.2.

### Acknowledgement

This work is partially supported by the Discovery Project grant of Australian Research Council.

### REFERENCES

- [1] K. J. Arrow, L. Hurwicz, H. Uzawa, *Studies in linear and nonlinear programming*, Stanford Mathematical Studies in the Social Sciences, vol. II. Stanford University Press, Stanford, 1958
- [2] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually  $o(k^{-2})$* , Preprint 2015.
- [3] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences*, 2 (2009), 183202.
- [4] S. Bonettini and V. Ruggiero, *On the convergence of primaldual hybrid gradient algorithms for total variation image restoration*, J Math Imaging Vis., 44 (2012), 236–253.
- [5] R. Bot and T. Hein, *Iterative regularization with a geeral penalty term—theory and applications to  $L^1$  and TV regularization*, Inverse Problems, 28 (2012), 104010.
- [6] H. Brakhage, *On ill-posed problems and the method of conjugate gradients*, In: Engl HW, Groetsch CW, editors. Inverse and ill-posed problems. Boston (MA): Academic Press 1987, 165–175.



- [7] L. M. Bregman, *The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys. 7 (1967), 200–217.
- [8] A. Chambolle and C. Dossai, *On the convergence of the iterates of the “fast iterative shrinkage/thresholding algorithm”*, J. Optim. Theory Appl., 166 (2015), 968–982.
- [9] A. De Cezaro, M. Haltmeier, A. Leito, O. Scherzer, *On steepest-descent-Kaczmarz methods for regularizing systems of nonlinear ill-posed equations*, Appl. Math. Comput., 202 (2008), no. 2, 596–607.
- [10] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [11] M. Haltmeier, A. Leitao and O. Scherzer, *Kaczmarz methods for regularizing nonlinear ill-posed equations I. Convergence analysis*, Inverse Probl. Imaging, 1 (2007), no. 2, 289–298.
- [12] M. Haltmeier, R. Kowar, A. Leitao and O. Scherzer, *Kaczmarz methods for regularizing nonlinear ill-posed equations II. Applications*, Inverse Probl. Imaging, 1 (2007), no. 3, 507–523.
- [13] M. Hanke, *Accelerated Landweber iterations for the solution of ill-posed equations*, Numer. Math., 60 (1991), 341–373.
- [14] M. Hanke, A. Neubauer and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), 21–37.
- [15] P. C. Hansen and M. Saxild-Hansen, *AIR Tools — a MATLAB package of algebraic iterative reconstruction methods*, J. Comput. Appl. Math., 236 (2012), 2167–2178.
- [16] M. Hegland, Q. Jin and W. Wang, *Accelerated Landweber iteration with convex penalty for linear inverse problems in Banach spaces*, Appl. Anal., 94 (2015), 524–547.
- [17] T. Hohage and F. Werner, *Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data*, Numer. Math., 123 (2013), 745–779.
- [18] K. Ito and B. Jin, *Inverse Problems: Tikhonov Theory and Algorithms*, World Scientific, Singapore, 2014.
- [19] Q. Jin, *Inexact Newton-Landweber iteration in Banach spaces with non-smooth convex penalty terms*, SIAM J. Numer. Anal., 53 (2015), no. 5, 2389–2413.
- [20] Q. Jin, *Inexact Newton-Landweber iteration for solving nonlinear inverse problems in Banach spaces*, Inverse Problems, 28 (2012), 065002 (15pp).
- [21] Q. Jin and U. Tautenhahn, *On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems*, Numer. Math., 111 (2009), 509–558.
- [22] Q. Jin and W. Wang, *Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals*, Inverse Problems, 29 (2013), 085011 (22pp).
- [23] Q. Jin and M. Zhong, *On the iteratively regularized Gauss-Newton method in Banach spaces with applications to parameter identification problems*, Numer. Math., 124 (2013), 647–683.
- [24] Q. Jin and M. Zhong, *Nonstationary iterated Tikhonov regularization in Banach spaces with uniformly convex penalty terms*, Numer. Math. 127 (2014), 485–513.
- [25] B. Kaltenbacher and B. Hofmann, *Convergence rates for the iteratively regularized Gauss-Newton method in Banach spaces*, Inverse Problems, 26 (2010), 035007 (21pp).
- [26] B. Kaltenbacher, A. Neubauer and O. Andreas, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, Radon Series on Computational and Applied Mathematics 6 Walter de Gruyter, Berlin, 2008.
- [27] R. Kowar and O. Scherzer, *Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems*, Ill-posed and inverse problems, 253–270, VSP, Zeist, 2002.
- [28] F. Natterer, *The Mathematics of Computerized Tomography*, SIAM, Philadelphia, 2001.
- [29] Y. Nesterov, *A method of solving a convex programming problem with convergence rate  $O(1/k^2)$* , Soviet Mathematics Doklady, 27 (1983), 372–376.
- [30] L. Rudin, S. Osher, and C. Fatemi, *Nonlinear total variation based noise removal algorithm*, Phys. D, 60 (1992), pp. 259–268.
- [31] F. Schöpfer, A. K. Louis and T. Schuster, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces*, Inverse problems, 22 (2006), 311–329.
- [32] F. Schöpfer and T. Schuster, *Fast regularizing sequential subspace optimization in Banach spaces*, Inverse Problems, 25(2009): 015013 (22pp).
- [33] T. Schuster, B. Kaltenbacher, B. Hofmann and K. S. Kazimierski, *Regularization Methods in Banach Spaces*, Radon Series on Computational and Applied Mathematics 10 Walter de Gruyter, Berlin 2012.

- [34] U. Tautenhahn and Q. Jin, *Tikhonov regularization and a posteriori rules for solving non-linear ill posed problems*, Inverse P roblems, 19 (2003), no. 1, 1–21.
- [35] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co., Inc., River Edge, New Jersey, 2002.
- [36] M. Zhu and T. F. Chan, *An efficient primaldual hybrid gradient algorithm for Total Variation image restoration*, CAM Report 08-34, UCLA (2008)

MATHEMATICAL SCIENCES INSTITUTE, AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, ACT  
2601, AUSTRALIA

*E-mail address:* `qinian.jin@anu.edu.au`